

search time (sec): 1030.430000

[illegible]

```

67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuProL 84
421 GAGCCGGAGATCTCAGTCCCTACAGTGCAGCTCCACCCGAGTCTKCHCC 470
84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100
471 TCCTCCCGGGCCGCCAGACATCCAGAGCCCTCTCCCGGGGGCCAG 520
101 GlyProArgProGlnThrSerProGlnThrLeuProLeuTyArgThrG 117
521 GTTCCAGACCCAGACTTCTCCAGAACCCCTCCCTTGTACAGGACACA 570
117 nSerArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyA 134
571 GNTCCGGGAAGGGTGGGCCACTTCGAGTCCCGCTCCACCTAGGGA 620
134 rgGluGluThrGlnGluLeuArgAlaAlaArgArgSerArgLeuArgasp 150
621 GAGAGGAGACCCAGGAGATTCGAGCGGCCAGGAGTCCCGGCTTCGAGAC 670
151 ProLeuLysProGlyMetPheGlyTyArgValArgValProPheAlaLeuPr 167
671 CCCATCAAGCCAGGATGTCGTTATGGGAGAGTCCCTTTGCATTGCC 720
167 oLeuHisArgAsnArgArgHisProArgSerProProArgSerGluLeuS 184
721 ACTGCACCGGAACCCAGCAGCAACCCCTCGGAGCCACCCAGATCTGAGCTGT 770
184 erLeuLeuSerSerArgGlyGluGluAlaLeuProSerProThrProArg 200
771 CCTGATCTCTCTAGAGGGGAGAGCCTATTCCTCCCTACTCCAGA 820
201 AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProTh 217
821 GCAGAGCCATTCTCGCAACGCGAGCCGCCCAACTGAGCTCCCTCCAC 870
217 rGluLeuSerValHis 222
871 AGAAGTGTCTGTCCAC 886

```

seq\_name: gb\_est2.BF536373

seq\_documentation\_block:

LOCUS BF536373 744 bp mRNA linear EST 11-DEC-2000  
 DEFINITION 602051728F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4191007 5',

mRNA sequence.

ACCESSION BF536373

VERSION BF536373.1 GI:11623741

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 744)

NH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: ILAM9519 row: g column: 08

High quality sequence stop: 685.

Location/Qualifiers

1..744

/organism="Mus musculus"

/strain="FVB/N"

FEATURES

source

/db\_xref="taxon:10090"  
 /clone\_image="4191007"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site\_1:  
 NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo  
 dr. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 139 a 231 c 208 g 166 t  
 ORIGIN

alignment\_scores:

Quality: 1093.00 Length: 259

Ratio: 4.836 Gaps: 2

Percent Similarity: 87.259 Percent Identity: 79.151

alignment\_block:

US-10-041-770-2 x BF536373 ..

Align seg 1/1 to: BF536373 from: 1 to: 744

```

428 ArgPheTyValArgHisThrGluLysValGlnAspGlyThrLeuCysG 444
|||||
2 CGATTCTATGTCCGACACACTGAAAGGTGCAGGATGGAACCTGTGTCA 51
444 nProGlyAlaProAspLeuCysValAlaGlyArgCysLeuSer.ProGly 460
|||||
52 CCTCGGATCCCTAGACATCTGTGGCTGGACGCTGCTGCGCTCCG... 98
461 CysAspGlyIleLeuGlySerGlyArgArgProAspGlyCysGlyValCy 477
|||||
99 .....GATGGCTGGGAGGCTG 115
477 sGlyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspA 494
|||||
116 CGGGGGTATGGTTTACCTGCCCGCTGGTTTCGGGAAATCTCACTGACC 165
494 rgGlyGlyProLeuGlyTyGlnLysIleLeuTrpIleProAlaGlyAla 510
|||||
166 GAGGGGGCCCTTGGCTATCAGAAGATCTTGTGGATCCTCGCGGGGGCC 215
511 LeuArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyIleAlaLe 527
|||||
216 TCCACCTTCACATTCCCACTTCGACCCAGTCCCAATTACCTCGCACT 265
527 uArgGlyProGlyArgSerIleIleAsnGlyAsnTrpAlaValAsp 544
|||||
266 CCGAGGGGCTGGGGCGCTCCATTATCAATGGGAAGTGGGCTGTGGATC 315
544 roProGlySerTyArgAlaGlyGlyThrValPheArgTyArgAsnArgPro 560
|||||
316 CTCAGGGTCTTATACAGCCATTGGGACTGCTTCCAGTATACCGTCTCT 365
561 ProArgGluGluGlyLysGlyGlySerLeuSerAlaGluGlyProThr 577
|||||
366 CCACGGGAAGCAAGGGGAGAGTCTGTACAGAGAAGCCCTATCCAC 415
577 rGlnProValAspValTyMetIlePheGlnGluGluAsnProGlyValP 594
|||||
416 TCAGCTGTGGACGCTCTATATGATCTTTCAGAGGACACACCCAGGTGTT 465
594 heTyGlnTyValIleSerSerProProIleLeuGluAsnProThr 610
|||||
466 TTTATCAGTATGTCTCTCTTCCCTCTCTGAGTCTAGAGAGTCTCTTCC 515
611 ProGluProProValProGlnLeuGlnProGluIleLeuArgValGluPr 627
|||||
516 ACAGAGCTCCGACCCCTTACGCTCAGCCTGAGATGCTGAGGGGGGAGCC 565
627 oProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnA 644
|||||
566 CCTACTCCCTCAGCCCCCGCCAGTCCGGGACACAGGACCCCTCCCAAC 615

```



```

644 rglnValArgIleProGlnMetProAlaProHisProArgThrPro 660
|||||
616 GTCAGGTGGGATCCCAAGTGCCTCCGACTCGTGTGAGGACGCC 665
|||||
661 LeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSe 677
|||||
666 ATGGGGTCTCTCTGCTGGTACTGGGAACAGTGGGGCACTCTGAGTGTTC 715
|||||
677 rAlaSerCysGlyLysGlyValTIP 685
|||||
716 ACATCCCTGTGGGCAAGGTGTGG 740
|||||

seq_name: gb_est2:BE808908
seq_documentation_block:
LOCUS BE808908 517 bp mRNA linear EST 25-APR-2001
DEFINITION 214215 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE808908
VERSION BE808908.1 GI:10240020
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Partes,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 70 row: N column: 6
Seq primer: ATTTAGTGACACTATAG.
FEATURES
source
1..517
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="vector: pcwv sport6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 84 a 175 c 179 g 79 t
ORIGIN

```

```

Align seg 1/1 to: BE808908 from: 1 to: 517
691 CysIleSerArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAl 707
|||||
1 TGCATTTCTCGAGAGTCAAGGAGGAGGAGTGGATGAACGACGCTGTGCCAT 50
|||||
707 aGlyAlaArgProAlaSerProGluProCysHisGlyThrProCysP 724
|||||
51 GGGTGCCAGGCCCCAGCCCTCCAGGAGCCCTGCCACGGCCCCCGGTGCC 100
|||||
724 roProTyrTrpGluAlaGlyLufTrpSerCysSerArgSerCysGly 740
|||||
101 CACCATACTGGAGCGCGGAGTGGAGTCTCTGCACCGCTTGTGTGGA 150
|||||
741 ProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGly 757
|||||
151 CCCGGCACCCAGCACCTCAGCTACGCTGCGGCGGAGGAGTTGGGGTGG 200
|||||
757 yGlySerSerValProGluArgCysGlyHisLeuProArgProAsnI 774
|||||
201 CGGCTCCTCAGTCCCTAGAGCGCTGGGGCACCTGCCCGACCCCAACA 250
|||||
774 leThrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluValGlySer 790
|||||
251 TCACCCAGCCCTGCCAGCTGGCCCTCTGTGCCCATTTGGAGGTTCTGCTCA 300
|||||
791 ProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerArgG 807
|||||
301 CCCGTGGAGTCACTCTCTGTGCGATGGGGCGGCTCAGAGGACCGGCA 350
|||||
807 nValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluCysA 824
|||||
351 GGTCCGCTGTGTCGCAACAACAGGGGATGAAGTGAAGTGAAGTGAAGTGC 400
|||||
824 laSerGlyProProGlnProProSerArgGluAlaCysAspMetGlyPro 840
|||||
401 CCTATGGCCCCCGGCCGCCAGAGAGAGGCTTGGCATTTGGGGCCC 450
|||||
841 CysThrThrAlaTrpPheHisSerAspTrpSerLysValSerProG 857
|||||
451 TGCACACGGCGCTGTTCCACAGCGACTGGAGCTCCAAGTCTCAGCGGA 500
|||||
857 u 857
501 G 501

seq_name: gb_est1:AW230557
seq_documentation_block:
LOCUS AW230557 495 bp mRNA linear EST 10-DEC-1999
DEFINITION uc65a11.y1 NCI_CGAP_Maml Mus musculus cDNA clone IMAGE:2647388 5'
similar to TR:060345 O60345 KIAA0605 PROTEIN.; mRNA sequence.
ACCESSION AW230557
VERSION AW230557.1 GI:6559853
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: uc65a11.x1
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

```

alignment_scores:
Quality: 904.00 Length: 167
Ratio: 5.722 Gaps: 0
Percent Similarity: 94.611 Percent Identity: 93.413
alignment_block:
US-10-041-770-2 x BE808908

```

www-bio.llnl.gov/bbrp/image/image.html

MGI:1027840  
Seq primer: -40RP from Gibco  
High quality sequence stop: 357.  
Location/Qualifiers  
1. 495  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2647388"  
/clone\_lib="NCI\_CGAP Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: Sall;  
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dr.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 84 a 155 c 149 g 107 t  
ORIGIN

alignment\_scores:  
Quality: 850.00 Length: 164  
Ratio: 5.484 Gaps: 0  
Percent Similarity: 94.512 Percent Identity: 87.805

alignment\_block:  
US-10-041-770-2 x AW230557 ..

Align seg 1/1 to: AW230557 from: 1 to: 495

647 ArgIleProGlnMetProAlaProHisProArgThrProLeuGlySe 663  
1 CGGATCCCAAGTGCCTCTCCGACTGCTGTCAGGACGACATGGGTC 50  
663 rProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerAlaSerC 680  
51 TTTCTGTGGATCTGAAACAAAGTGGGGCACTGTGAGTTTTCAGCATCT 100  
680 ySGLYLysGlyValTrpArgProIlePheLeuCysIleSerArgGluSer 696  
101 GTGGCAAGGTGTTGGCACCCCATTTCTCTGCATTTCCTGGAGTCA 150  
697 GlyGluGluLeuaspGluArgSerCysAlaAlaGlyAlaArgProAla 713  
151 GGAGAGGAGTTGGATGAACAGAGCTGTGTGGGGCCAGACCCCCAGC 200  
713 aSerProGluProCysHisGlyThrProCysProProTyrTrpGluAlaG 730  
201 TTCCCTGTACCTGCCATGGACCCCGGTCTCTCCATCTGGAGGCTG 250  
730 lyGluTrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisArg 746  
251 GCAGTGGAGATCTCTGCAGCCGATCTGTGGCCCTGGCACCCAGCCGC 300  
747 GlnLeuGlnCysArgGlnGluPheGlyGlyGlySerValProPr 763  
301 CAGCTACTCTGCAGACAGGATTCGGAGGTGGGGCTCTCGGTACTCTC 350  
763 oGluArgCysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnL 780  
351 AGAGCGCTGTGGACATCTCCCGCGGCCAACATCACCCAGCCTTGTGAGC 400  
780 euArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSer 796  
401 TGACCTCTGTGGCCACTGGAGATAGTCTCCCGCGGAGCCAGTGTCT 450  
797 ValArgCysGlyArgGlyGlnArgSerArgGlnValArgCys 810  
451 GTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 492

seq\_name: gb\_est2:BI330332

seq\_documentation\_block:  
BI330332 866 bp mRNA linear EST 30-JUL-2001  
LOCUS 602983024F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5135860 5',  
DEFINITION mRNA sequence.  
ACCESSION BI330332  
VERSION BI330332.1 GI:15014989  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 866)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-f@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM11333 row: p column: 05  
High quality sequence start: 3  
High quality sequence stop: 548.  
Location/Qualifiers  
1..866  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5135860"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site: 2: Sall; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 149 a 332 c 237 g 148 t  
ORIGIN

alignment\_scores:  
Quality: 848.50 Length: 286  
Ratio: 3.839 Gaps: 13  
Percent Similarity: 77.273 Percent Identity: 67.483

alignment\_block:  
US-10-041-770-2 x BI330332 ..

Align seg 1/1 to: BI330332 from: 1 to: 866

376 GlnAlaProCysProGluGlnProAspProArgAlaLeuGlnCysAl 392  
2 CAGGAGCCTTGCCCCCTGGAGCATGCCAGACCCAGGCCCTACAGTGTGC 51  
392 aAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrGlnTrpGluProp 409  
52 CCCTTTTGTACTCCAGAGATTTCATGGCCAGCTGTACCAGTGGAGCCCT 101  
409 heThrGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProArg 425  
102 TCACCGAGTTTCAGGGCTCCAGCCCTGTGACTGACTGACTGCGCCCGCT 151  
426 GlyPheArgPheTyrValArgHisThrGluLysValGlnAspGlyThrLe 442  
152 GCGTTCCTCCGATTTCTATGTCGACACACTGAAAGGTGCAGATGGAACCT 201  
442 uCysGlnProGlyAlaProAspIleCysValAlaGlyArgCysLeuSerP 459

451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@uiowa.edu  
Oligo-dt track not found, Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=NO. Location/Qualifiers  
1. .592  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DY0-ckr-1-07-0-UI"  
/clone\_lib="UI-R-DY0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not 1; Site\_2: Eco RI; UI-R-DY0 is a  
non-normalized Rat cartilage library (RC) constructed in  
pT37 PAC vector according to the procedure described by  
Bonaldi, Lennon & Soares (Genome Research 6:  
791-806, 1996). The oligonucleotide used to prime first  
strand synthesis contained the sequence tag CTAATGGAGC  
between the Not I cloning site and d18 stretch. The Rat  
cartilage tissue was provided by Dr Jeff Stevens at the  
University of Iowa.  
TAG\_SEQ=None found"

BASE COUNT 110 a 181 c 192 g 108 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 828.00 Length: 160  
Ratio: 5.483 Gaps: 1  
Percent Similarity: 94.375 Percent Identity: 88.750  
alignment\_block:  
US-10-041-770-2 x BM391602 ..  
Align seg 1/1 to: BM391602 from: 1 to: 592  
699 GluLeuAspGluArgSerCysAlaAla GlyAlaArgProAlaSerP 715  
27 CAGTTGGATGAACAGAGCTGTCCTGGCCGAGACCCAGCTCC 76  
715 rogluProCysHisGlyThrProCysProProTyrTrpGluAlaGlyGlu 731  
77 CTGAATCTGTCATAGACACCGTGTCCCTACTGGAGGCTGGCGAG 126  
732 TrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLe 748  
127 TGGAGCTCTGTCAGCGCTCTCTGTGGCCCTGGCCAGCATCGCCAGCT 176  
748 uGlnCysArgGlnGluPheGlyGlyGlySerSerValProGluA 765  
177 GCTCTGCAGACAGGAATTCGGAGCGGTGGTCTCTCAGTGCCTCCAGAGC 226  
765 rgCysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArg 781  
227 GTTGGGACATCTCCCGCGCCCAACATCACCCAGTCTGTGAGCTGGC 276  
782 LeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSerValar 798  
277 CTTTGGCCACTGGAGATTAGTCCCTGGAGCCAGTCTGTGTGCG 326  
798 gCysGlyArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnG 815  
327 CAGTGTCTGTGTGTCAGAGGAGCGCGCAAGTGGGTGTGTGTGAAGTAATG 376  
815 lyAspGluValSerGluGlnGluCysAlaSerGlyProProGlnProPro 831

202 GTGTACGCTGTAGATCCCTAGACATCTGTGTGCTGGAGCCTGCTGAGCC 251  
459 rogluCysAspGlyLeuLeuGlySerGlyArgPro. AspGlyCysG1 475  
252 CTGGCTGTGATGGGTCCTGTGCTGCGAGCGTCCGGCATGGCTGCGG 301  
475 yValCysGlyGlyAsp. AspSerThrCysArgLeuValSerGly. AsnLe 491  
302 AGTCTGCGGGGTGACTGTCTTACCTGCGGCTGGTTCGGGCAAACT 351  
491 uThrAspArgGlyGlyProLeuGly. TyrGln. LysIleLeuTrpIlePr 507  
352 CACTGACCGAGGGGCCCTTGGGCTATCAGATGATCTGTGGATCC 401  
507 oAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTr 524  
402 TGGCGGGGCTCCACCTTCACATTCCAGTTCGAGCCAGT. TCCAAT 450  
524 yLeuAlaLeuArgGlyProGlyGlyArgSerIle. IleAsnGlyAsnTr 540  
451 ACCTGCACTCGAGGGCGTGGGGCGCTCCATTCATCAGTGGGAAC 500  
540 pAlaValAspProPro. GlySerTyrArgAlaGlyGlyThrValPheArg 556  
501 GGCTGTGATCTCCAGGCTCTCTACAGCCATCGCGACTGTCTTCCAG 550  
557 TyrAsnArgProArg. GluGlu... GlyLysGlyGluSerLeuSerA 572  
551 TATAAGCTCTCCACGGGCAAGACAGGCGCGGAGGAGTCTGTCTCAG 600  
572 laGluGlyProThrThrGlnProVal. AspValTyrMetIlePheGlnG1 588  
601 CCAAGGCTTACCACTACGCTGCGCGCTCTCTCGCACCCACCCAC 650  
588 uGluAsnProGlyValPheTyrGlnTyrValIleSerSerProProI 605  
651 CGAGGACCCCGAGGTTGCGCGGACCCGCGCATCTCATCCACCC 700  
605 leLeuGluAsnProThrProGluPro..... 613  
701 TCCCGGCGAGCGGAGAGAGGCGCTCCACAGACGCTCCGACCCAT 750  
614 .....ProValProGlnLeuGlnProGluIleLeuArgValGluProPr 628  
751 GTAGCCCCCAGGCGCGAGTGTG... CCGAGGGCGGCAACCAACAGCC 797  
628 oLeuAlaProAlaProArgProAlaArgThrProGlyThr 641  
798 CACAGCCCGCGCCACCGCGCC...ACACGACCAAG 834  
seq\_name: gb\_est2:BM391602  
seq\_documentation\_block:  
LOCUS BM391602 592 bp mRNA linear EST 17-JAN-2002  
DEFINITION UI-R-DY0-ckr-1-07-0-UI.s1 UI-R-DY0 Rattus norvegicus cDNA clone  
UI-R-DY0-ckr-1-07-0-UI 3', mRNA sequence.  
ACCESSION BM391602  
VERSION BM391602.1 GI:18191655  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 592)  
AUTHORS Bonaldi,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa

```

377 GCCATGAGTGGGCAAGCAGGAGTGTCTCGGGCCCCGCCACCTCCC 426
|||||
832 SerArgGluAlaCysAspMetGlyProCysThrThrAlaTyrPheHis 848
|||||
427 AGCAGAGAGCCCTGTACATGGCCCTTACACAGCCTGTTCTTACAG 476
|||||
848 rAspTyrSerSerLysValSerProGlu 857
|||||
477 TGACTGGAGTTCGAAGTGTCTCAGCCGAG 504
|||||
```

```

seq_name: gb_est2:BI094238
seq_documentation_block:
LOCUS      BI094238               865 bp      mRNA      linear      EST 20-JUN-2001
DEFINITION 602860118F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5001173 5',
            mRNA sequence.
ACCESSION  BI094238
VERSION    BI094238.1 GI:14512568
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 865)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LAM11034 row: d column: 06
            High quality sequence stop: 510.
```

```

FEATURES
            Location/Qualifiers
                1..865
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5001173"
                    /clone_lib="NIH_MGC_10"
                    /cell_line="MGC36"
                    /lab_host="DH10B"
                    /note="organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                    Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                    Average insert size 1.5 kb. Library prepared by Life
                    Technologies."
BASE COUNT      240 a 187 c 303 g 135 t
ORIGIN
```

```

alignment_scores:
    Quality: 808.50      Length: 204
    Ratio: 4.620        Gaps: 6
    Percent Similarity: 85.784      Percent Identity: 82.843
alignment_block:
US-10-041-770-2 x BI094238 ..
Align seg 1/1 to: BI094238 from: 1 to: 865
425 ArgGlyPheArgPheTyrValArgHsThrGluLysValGlnAspGlyTh 441
|||||
1 CGTGGCTTCGGCTCTATGTCCGCTCACACATGAAGAAGTCCAGGATGGGAC 50
|||||
441 rLeuGlyLnpProGlyAlaProAspIleCysValAlaGlyArgCysLeus 458
|||||
51 CCTGTGTGAGGTGGAGCCCTGTGATCTGTGTGGCTGGAGCGTGTCTGA 100
|||||
```

```

458 erProGlyCysAspGlyIleLeuGlySerGlyArgArgProAspGly.Cy 474
|||||
101 GCCCGGGCTGTGATGGGATCCTTGGCTCTGGCAGGGCTCTCTGATGGCTTG 150
|||||
474 sglyValCysGlyGlyAspSerThrCysArgLeuValSerGlyAsnL 491
|||||
151 TGGAGTCTGTGGGGTGTATCTTACTCTGCGCTTGTTCGGGAACC 200
|||||
491 euThrAspArgGlyGlyProLeuGlyTyrGlnLysIleLeuTyrPilePro 507
|||||
201 TCACTGACCGAGGGGGCCCTGGGCTATCAGAAGATCTTGTGATTCCA 250
|||||
508 alaglyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTy 524
|||||
251 GCGGGAGACCTTGGGCTCCAGATTGCCAGCTCCGGCTAGCTCCAATA 300
|||||
524 rLeuAlaLeuArgGlyProGlyArg.SerIleIleAsnGlyAsn... 539
|||||
301 CTGGCACTTGTGGCCCTGGGGCTGTCTCATCATCAATGGGAACCTG 350
|||||
540 rtpAlaValAspProProGlySerTyrArgAlaGlyGlyThrValPheAr 556
|||||
351 TGGCTGTGGATCCCGCTGGGCTCTACAGGCGCGGGACCGCTTTCG 400
|||||
556 gTyrAsnArgProProArgGlu.GluGlyLysGlyGlySerLeuSerAla 572
|||||
401 ATATAACCGTCTCTCCAGGAAGGAGGCAAAAGGGAGAGTCTGTGCGCT 450
|||||
573 GluGlyProThrThrGlnProValAspValTyrMetIlePheGlnGlu1 589
|||||
451 GAAGGCCCAACAC.CAGCGTGTGATGTCTATATGATCTTTCAGGAGGA 499
|||||
589 uAsnProGlyValPheTyrGlnTyrValIleSerSerProProPile. 605
|||||
500 AAACCCAGCGGTAGATAATCGAGTAGTGTCTGAGTCTGTCTCACGCTCG 549
|||||
606 LeuGluAsnPro.....ThrProGluProProValPr 616
|||||
550 CTCGCAATCTCTGGAGAAAGCGCAAGACACAGAGCGCGCCCTAGAGC 599
|||||
616 ogIn 617
|||||
600 GCAA 603
```

```

seq_name: gb_est1:AW763191
seq_documentation_block:
LOCUS      AW763191               513 bp      mRNA      linear      EST 04-MAY-2000
DEFINITION ur-60h06.y1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:3154715 5',
            similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ;, mRNA sequence.
ACCESSION  AW763191
VERSION    AW763191.1 GI:7695127
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 513)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
MGI:1057471
```

```
seq_documentation_block:
LOCUS      BM148104          437 bp      mRNA      linear      EST 30-NOV-2001
DEFINITION TCAAP:Q8352 pediatric acute myelogenous leukemia cell (FAB M1)

```

```

116 GCACGGCCCCGAGAGTCTCTGGGACCTTGGGTCCAGTGGGCTCTTGC 165
57 SerGlnProCysGlyValGlyValGlnArgSerArgThrCysGlnLe 73
166 TCCACGCTCTGGGGTGGGGTGGAGCGCAGGAGCGGACATGTACACT 215
73 uProThrValGlnLeuHisProSerLeuProLeuProProProPro 90
216 CCTACAGTGCAGCTCCACCGGAGTCTGCCCTCCCTCCCGGGCCCCAA 265
90 rgHisProGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThr 106
266 GACATCCAGAACCTCTCTCCCGGGGCCAGGTCACAGACCCAGACT 315
107 SerProGluThrLeuProLeuThrArgThrGlnSerArgGlyArgGly 123
316 TCTCAGAAACCTCCCTCTGTACAGGACACAGTCTCGGGGAGGGTGG 365
123 yProLeuArgGlyProAlaSerHisLeuGlyValGlnGluThrGlnGlu 140
366 CCACCTTCAAGGACCGCTGTCACCTAGGGAGAGGAGAGCCAGGAGA 415
140 leArgAlaAlaArgArg 145
416 TTCCAGCGGACAGCAGG 432

```

seq\_name: gb\_est1:AI683106

seq\_documentation\_block:

LOCUS AI683106 597 bp mRNA linear EST 16-DEC-1999  
DEFINITION tx01d11.x1 NCI-CGAP\_Ut4 Homo sapiens cDNA clone IMAGE:2267925 3',  
similar to TR:002661 002661 SCO-SPONDIN ; contains MSRI.t2 MSRI  
repetitive element ; , mRNA sequence.

ACCESSION AI683106

VERSION AI683106.1 GI:4893278

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 597) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1132 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 369.

Location/Qualifiers

FEATURES

source

1. 597

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2267925"

/clone\_lib="NCI-CGAP\_Ut4"

/tissue\_type="serous papillary carcinoma, high grade, 2

pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.48 kb. Life Technologies catalog #:

11542-016"

BASE COUNT 118 a 162 c 180 g 136 t 1 others

ORIGIN

alignment\_scores:

Quality: 729.00 Length: 175

Ratio: 4.796 Gaps: 0

Percent Similarity: 86.857 Percent Identity: 78.286

alignment\_block:

US-10-041-770-2 x AI683106/rev ..

Align seg 1/1 to reverse of: AI683106 from: 1 to: 597

```

45 GlyValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysGI 61
|||||
597 GGTGTTTGGGGCCCTTGGGTCAAGTGGGCTCTTGTCTCCAGAGCCCTCGG 548
61 yValGlyValGlnArgSerArgThrCysGlnLeuProThrValGlnL 78
|||||
547 GGTGGGGTGCAGCGCAGGAGCGCAACATGTACAGTCTCTTACAGTGGAGC 498
78 euHisProSerLeuProLeuProProArgProProArgHisProGluAla 94
|||||
497 TCCACCGGAGTTTGGCCCTCCCTCCCGGGCCCCCAGACATCCAGAGCC 448
95 LeuLeuProArgGlyGlnGlyProArgProGlnThrSerProGluThrLe 111
|||||
447 TTCCTCCCGGGGCCAGGCTCCAGACCCAGATTTCCTCCAGAAACCT 398
111 uProLeuThrArgThrGlnSerArgGlyArgGlyGlyProLeuArgGlyP 128
|||||
397 CCCTTTGTACAGGACCCAGATTTCGGGGAAGGGTGGCCCTTTTCGGGGTC 348
128 roAlaSerHisLeuGlyValGlyArgGluThrGlnGluLeuArgAlaAla 144
|||||
347 CCGTTCCCTTAGGGAGAGAGGAGCCCGCAGAGTTTCGAGCGGCCAGG 298
145 ArgSerArgLeuArgAspProIleLysProGlyMetPheGlyTyrGlyAr 161
|||||
297 GGGTCCCGGTTTCAGACCCCATCAGCCAGCAATGTGGTTATGGGAG 248
161 gValProPheAlaLeuProLeuHisArgAsnArgHisProArgSerP 178
|||||
247 AGTGCCTTTCATGCCANTGTCACGGGACCGCGGGCCCTCGGGGCC 198
178 roProArgSerGluLeuSerLeuLeuSerSerArgGlyGluGluAlaLe 194
|||||
197 CCCCAGATTTTGGAGCTGTCCCTGTTTTTTTAGGGGGGAGGGGTTTTT 148
195 ProSerProThrProArgAlaGluProPheSerAlaAsnGlySerProGI 211
|||||
147 CGGTCCCTTACTCCAGAGCAGAGCCCTTTTCCGCAAGGCGGCCCCCA 98
211 nThrGluLeuProProThrGluLeu 219
|||||
97 AATTGGGCTCCTCCCGCAGAACTG 73

```

seq\_name: gb\_est2:BG438243

seq\_documentation\_block:

LOCUS BG438243 625 bp mRNA linear EST 14-MAR-2001

DEFINITION 602490152F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:462245 5',

mRNA sequence.

ACCESSION BG438243

VERSION BG438243.1 GI:13344749

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 625) Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NTH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)



## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: crapbe-remail.nih.gov  
 Tissue Procurement: DCTD/DTP/Gazdar  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLCMI386 row: g column: 14  
 High quality sequence stop: 406.

## FEATURES

source  
 1. .625

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4622245"  
 /clone\_lib="NIH\_MGC\_18"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pORF7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dr priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GCCACGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

BASE COUNT 188 a 161 c 174 g 102 t

## ORIGIN

## alignment\_scores:

Quality: 681.00 Length: 180  
 Ratio: 4.540 Gaps: 6  
 Percent Similarity: 83.333 Percent Identity: 81.667

## alignment\_block:

US-10-041-770-2 x BG438243 ..

Align seg 1/1 to: BG438243 from: 1 to: 625

473 GlyCysGlyValCysGlyGlyAspSerThrCysArgLeuValSer G 489  
 |||||  
 3 GGCTGTGGAGTCGTGGGGTGATGATCTACCTGTCGCTGTTCTGG 52  
 489 lYAsnLeuThrAspArgGlyGlyProLeuGlyTyrClnLysIleLeuTyr 505  
 |||||  
 53 GGAACCTCACTGACCGAGGGGGCCCTGGGCTATCAGAAGATCTTGTGG 102  
 506 IleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSe 522  
 |||||  
 103 ATTCAGCGGGAGCCCTGGGCTCCAGATTGCCAGCTCGGGCTAGCTC 152  
 522 rAsnTyrLeuAlaLeuArgGlyProGlyArgSerIleIleAsnGlyA 539  
 |||||  
 153 CAACCTACCTGGCCTCGTGGCCCTGGGGCCGGTCCATCATCAATGGGA 202  
 539 snTyrAlaValAspProProGlySerTyrArgAlaGlyGlyThrValPhe 555  
 |||||  
 203 ACTGGGCTGGATCCCTGGGTCTCTACAGGGCCGGGACCGCTTT 252  
 556 ArgTyrAsnArgProProArg. GluGlyGlyLysGlyGluSerLeuSerA 572  
 |||||  
 253 CGATATAACCGTCTCCAGGTGGAGGGCAAGAGGAGTCTGTGCG 302  
 572 lagGlyGlyProThrGlnProValAspValTyrMetIlePheGlnGlu 588  
 |||||  
 303 CTGAAGCCCCACACACCGCCCTGGGTGCTATATGATCTTCAGAG 352  
 589 GluAsnProGlyValPheTyrGlnTyrValIleSerSerProProIle 605  
 |||||  
 353 GAAACCCAGGCGT.TTTTATCATGATGATCTCTTTCACCTCTCCAAAT 401

605 eLeu...GluAsnPro.ThrProGlu...ProProValProGlnLeuGln 619  
 |||||  
 402 CTTTGAGAAACCGAACACACGAGCGCGCTGTACCCACACAAGAT 451  
 |||||  
 620 ProGluIleLeuArgValGluProProLeuAlaProAlaProArgProAl 636  
 |||||  
 452 ACAGAA.....CGAGCAGA 465  
 |||||  
 636 aArgThrProGlyThrLeuGlnArgGlnValArg 647  
 |||||  
 466 GTCCTGGAGGAGCGACGACCAACGCCACACGCA 499

seq\_name: gb\_est2:BF993048

## seq\_documentation\_block:

LOCUS BF993048 360 bp mRNA linear EST 23-JAN-2001  
 DEFINITION IL5-GN0178-311000-204-f10 GN0178 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF993048  
 VERSION BF993048.1 GI:12399371  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 360)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

## TITLE

sequence tags

## JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## MEDLINE

20202663

## COMMENT

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t3=IL5-GN0178-311000-204-f10&t3=2000-10-31&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 305.

## FEATURES

## source

1. .360  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="GN0178"  
 /dev\_stage="Adult"  
 /note="Organ: placenta\_normal; Vector: puc18; Site\_1: SmaI  
 ; Site\_2: SmaI; A mini-library was made by cloning  
 products derived from ORESTES PCR (U.S. Letters Patent  
 Application No. 196,716 - Ludwig Institute for Cancer  
 Research) profiles into the pUC 18 vector. Reverse  
 transcription of tissue mRNA and cDNA amplification were  
 performed under low stringency conditions."  
 59 a 80 c 142 g 78 t

## BASE COUNT

## ORIGIN

## alignment\_scores:

Quality: 625.00 Length: 119  
 Ratio: 5.342 Gaps: 0  
 Percent Similarity: 98.319 Percent Identity: 98.319

## alignment\_block:

US-10-041-770-2 x BF993048/rev ..

Align seg 1/1 to: BE757541 from: 1 to: 571

745 HisArgGlnLeuGlnCysArgGlnGluPheGlyGlyGlyGlySerSerVa 761  
|||||  
3 CACGCTCAGCTACGCTCCGCGCAGAGATTGGGGGTGGGGCTCTCACT 52  
|||||  
761 lProPGluAArgCysGlyHisLeuProAArgProAsnIleThrGlnSerC 778  
|||||  
53 GCCCCTAGAGGGCTGGGGGACCTGCCCCGACCCACATCACCAGCCCT 102  
|||||  
778 ysGlnLeuA-gLeuCysGlyHisTrpGluValGlySerProTrpSerGln 794  
|||||

**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

REFERENCE  
AUTHORS  
1 (bases 1 to 370)  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
Dias Neto, E., Garcia Correa, P., Verjowski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, P., Matsukuma, A., Baia, G.S., Simpson, D.H., Brustein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMI&td=PMI-GN0180-  
241106-003-e05&t3=2000-11-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 370.

FEATURES  
source  
1..370  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0180"  
/dev\_stage="Adult"  
/note="Organ: Placenta\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT  
ORIGIN  
80 a 127 c 98 g 64 t 1 others

alignment\_scores:  
Quality: 594.00 Length: 117  
Ratio: 5.165 Gaps: 1  
Percent Similarity: 98.291 Percent Identity: 98.291

alignment\_block:  
US-10-041-770-2 x BG005397 ..

Align seg 1/1 to: BG005397 from: 1 to: 370

108 ProGluThrLeuProLeuTyr.ArgThrGlnSerArgGlyArgGlyGlyP 124  
|||||  
21 CCAGAAACCCCTCCCTNGTACAGGACAGTCTCGGGAGGGGTGGCC 70  
|||||  
124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluIle 140  
|||||  
71 CACTTCGAGTCCCGCTTCCACCTAGGAGAGAGAGAGAGAGAGAT 120  
|||||  
141 ArgAlaAlaArgSerArgLeuArgAspProIleLysProGlyMetPh 157  
|||||  
121 CGAGCGCCAGGAGGTCCCGGCTTCGAGACCCCATCAAGCCAGGAATGT 170  
|||||  
157 eGlyTyrGlyArgValProPheAlaLeuProLeuHisArgAsnArgH 174  
|||||  
171 CGGTTATGGAGAGTCCCTTTTCATGTGCACACCGACCGAGGCG 220  
|||||  
174 isProArgSerProProArgSerGluLeuSerLeuIleSerSerArgGly 190  
|||||  
221 ACCCTCGAGCCACCCAGATCTGAGCTGTCCCTGATCTCTTCTAGAGG 270  
|||||  
191 GluGluAlaIleProSerProThrProArgAlaGluProPheSerAlaAs 207  
|||||  
271 GAAGAGGCTATTCCGTCCTTACTCAAGAGCAGAGCCATTCCTCCCAAA 320  
|||||  
207 nGlySerProGlnThrGluLeuProProThrGluLeuSerValHisThr 223  
|||||  
321 CGGACGCCCCCAACATGAGCTCCCTCCACAGAACTGTCTGTCCACACC 369  
|||||

seq\_name: gb\_est2:BG900891

seq\_documentation\_block:  
LOCUS BG900891 449 bp mRNA linear EST 06-NOV-2001  
DEFINITION HOA7-1-H11 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,  
mRNA sequence.  
ACCESSION BG900891  
VERSION BG900891.1 GI:14311140  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 449)  
AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,  
Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and  
Lark,M.W.  
IDENTIFICATION AND INITIAL CHARACTERIZATION OF 5000 EXPRESSED  
SEQUENCED TAGS (ESTs) EACH FROM ADULT HUMAN NORMAL AND  
OSTEOARTHRITIC CARTILAGE CDNA LIBRARIES  
Osteoarthritic Cartilage 9 (7), 641-653 (2001)  
21482651  
CONTACT: Sanjay Kumar  
UN2109  
GlaXosmithKline  
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
Tel: 610-270-7245  
Fax: 610-270-5598  
Email: sanjay.kumar-1@gsk.com  
Seq primer: T7  
Location/Qualifiers  
1..449  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HOA (Human Osteoarthritic Cartilage)"  
/tissue\_type="cartilage"  
/lab\_host="E.coli DH10 B"  
/note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;  
Directional"  
BASE COUNT 71 a 158 c 134 g 86 t  
ORIGIN

alignment\_scores:  
Quality: 579.00 Length: 106  
Ratio: 5.567 Gaps: 0  
Percent Similarity: 98.113 Percent Identity: 96.226

alignment\_block:  
US-10-041-770-2 x BG900891 ..

Align seg 1/1 to: BG900891 from: 1 to: 449

1 MetGluAsnTrpThrGlyArgProTrpLeuTyrLeuLeuLeuLeuLeu 17  
|||||  
132 ATGGAGAACTGGACTGGCAGGCCCTGGCTGTATCTGCTGCTCTCTGTC 181  
|||||  
17 rleuProGlnLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuG 34  
|||||  
182 COTCCCTCAGCTCTCTTGGATCAGAGGTGTGTGCCACACACTCTCTTC 231  
|||||  
34 InThrProThrGluGlyGlnGlyProGluGlyValTrpGlyProTrp 50  
|||||  
232 AGACACCTACAGAGAGGCCAGGCCCGAAGGTGTCTGGGACCTTGG 281  
|||||  
51 ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgAr 67  
|||||  
282 GTCCAGTGGGCTCTTGTCCAGCCCTCGGGGTGGGGTGCAGCGCAA 331  
|||||  
67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuProL 84  
|||||  
332 GAGCCGGACATGTAGTCCCTCCTACAGTGCAGCTTCACCCGAGTCTGCC 381  
|||||  
84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100  
|||||



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:40:30 ; Search time 292.16 Seconds  
(without alignments)  
15479.027 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atggagaactgactggcag.....octcagccttcacgcatag 2634

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1259	47.8	2680	22 AAC91178	ADAM gene #2. Hom
2	472.4	17.9	1020	22 AAD16773	Human novel protei
3	472.4	17.9	1998	22 AAD16758	Human novel protei
4	165.2	6.3	2964	22 AAH16636	Human cdna sequenc
5	118.2	4.5	704	22 AAH99398	Human protein enco
6	113.2	4.3	2805	22 AAS97176	Human metalloprote
7	106.8	4.1	3675	24 AAS97174	Human metalloprote
8	98.2	3.7	5338	22 AAH49372	Human metalloprote
9	97	3.7	9295	23 ABL02495	Drosophila melanog

10	97	3.7	21314	23 ABL02494	Drosophila melanog
11	96.6	3.7	5061	22 AAF82149	Human ADAM type me
12	96.2	3.7	5353	22 AAD04475	Human 27875 ADAM-T
13	95.6	3.6	2274	22 AAD07967	Human ADAM type me
14	95.6	3.6	2727	22 AAD07956	Human ADAM type me
15	95.6	3.6	2879	22 AAH49370	Human metalloprote
16	95.6	3.6	3160	22 AAD07968	Human metalloprote
17	95	3.6	3533	22 AAD04494	Human ADAM type me
18	94	3.6	2848	22 AAF63439	Human ADAM type me
19	92	3.5	3776	24 AAH43769	ADAMTS-E DNA. Hom
20	91.6	3.5	3312	22 AAF82157	Human ADAM type me
21	91.6	3.5	3312	24 AAS97173	Human metalloprote
22	91.4	3.5	3199	22 AAS02432	Human ADAM type me
23	89.8	3.4	2875	22 AAD17501	Human ADAMTS-J1.4
24	89.4	3.4	814	22 AAH08271	Human ADAM type me
25	88.4	3.4	3409	22 AAF63448	Human ADAMTS-10 al
26	85.4	3.2	6505	21 AAF75588	Human OREFX ORF143
27	80.6	3.1	4287	24 AAS97177	Human metalloprote
28	77.4	2.9	1119	22 AAD14366	Novel human protei
29	77.4	2.9	1953	22 AAD14368	Novel human protei
30	77.4	2.9	2316	22 AAD14370	Novel human protei
31	77.4	2.9	4854	22 AAD14371	Novel human protei
32	74.2	2.8	4067	21 AAA95823	Human metalloprote
33	74	2.8	3218	22 AAF63440	Human ADAMTS-7 CDN
34	73.6	2.8	1698	22 AAD17500	Human ADAMTS-J1.3
35	73.6	2.8	1966	22 AAD17502	Human ADAMTS-J1.4
36	73.4	2.8	3263	22 AAH49371	Human metalloprote
37	73.4	2.8	3885	22 AAF63445	Human ADAMTS-10 CD
38	73	2.8	846	22 ABA08559	Human HSP124 pro
39	73	2.8	1341	22 AAD14365	Novel human protei
40	73	2.8	2175	22 AAD14367	Novel human protei
41	73	2.8	2538	22 AAD14369	Novel human protei
42	73	2.8	5076	22 AAD14364	Novel human protei
43	73	2.8	7260	22 AAD17757	Human novel KIRA12
44	73	2.8	8578	22 AAD14372	Novel human protei
45	72.6	2.8	1905	22 AAD14799	Human ADAMTS-J1.2

ALIGNMENTS

RESULT 1

AAC91178

ID AAC91178 standard; DNA; 2680 BP.

XX

AC AAC91178;

XX

DT 20-MAR-2001 (first entry)

XX

DE ADAM gene #2.

XX

KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;

KW ovarian; ss.

XX

OS Homo sapiens.

XX

PN WC200073323-A2.

XX

PD 07-DEC-2000.

XX

XX

XX 25-MAY-2000; 2000WO-US14308.

XX

PR 27-MAY-1999; 99US-0136388.

PR 09-JUL-1999; 99US-0142930.

PR 28-JAN-2000; 2000US-0178717.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;

XX

DR - WPI; 2001-016507/02.

XX

PT Seven nucleic acid molecules encoding ADAM polypeptides containing a

A A

PT disintegrin and metalloprotease domain, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -

XX Claim 1; Page 266; 287pp; English.

CC The present invention relates to seven members of the ADAMs (proteins  
 CC which contain a Disintegrin And Metalloprotease domain) protein family.  
 CC The ADAMs proteins and DNA may be used to treat disease, as a food  
 CC additive or preservative, for chromosome identification, as probes  
 CC for diagnosing a disorder related to the female reproductive system,  
 CC particularly breast and/or ovary cancer. They are also useful in the  
 CC gene therapy of breast and ovarian cancer.

XX Sequence 2680 BP; 510 A; 864 C; 767 G; 537 T; 2 other;

Query Match 47.8%; Score 1259; DB 22; Length 2680;  
 Best Local Similarity 100.0%; Pred. No. 8.9e-287;  
 Matches 1259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 ctgaaagggtccaggatgggacccctgtgtcagccttgaggcccttgacatctgtgtggtg 1360  
 DB 3 ctgaaagggtccaggatgggacccctgtgtcagccttgaggcccttgacatctgtgtggtg 62  
 QY 1361 gacgctgtcagcccggtgtgatggatccttggctctgagcgctcctgatggct 1420  
 DB 63 gacgctgtcagcccggtgtgatggatccttggctctgagcgctcctgatggct 122  
 QY 1421 gtggagctgtgggggtgatattacattcctgtcgcctgttttcgggaaacctcaactgacc 1480  
 DB 123 gtggagctgtgggggtgatattacattcctgtcgcctgttttcgggaaacctcaactgacc 182  
 QY 1481 gaggggggcccccctgggtatcaagaatcttggattccagcgggagccttcggtcc 1540  
 DB 183 gaggggggcccccctgggtatcaagaatcttggattccagcgggagccttcggtcc 242  
 QY 1541 agattgccaggtccggtcctagtcacactacattcctggtccttgccttggccctggggcggt 1600  
 DB 243 agattgccaggtccggtcctagtcacactacattcctggtccttgccttggccctggggcggt 302  
 QY 1601 ccattcatcaatgggaactgggtcgtggtatccccctgggtcctacagggcgcgagaccg 1660  
 DB 303 ccattcatcaatgggaactgggtcgtggtatccccctgggtcctacagggcgcgagaccg 362  
 QY 1661 tctttcgatataacgctcctccaggaggaggggaaaggagagtgctgtcgccggaag 1720  
 DB 363 tctttcgatataacgctcctccaggaggaggggaaaggagagtgctgtcgccggaag 422  
 QY 1721 gccccaccaccagcctgtggtatgtatattcatttcaggaggaaaccccaagggttt 1780  
 DB 423 gccccaccaccagcctgtggtatgtatattcatttcaggaggaaaccccaagggttt 482  
 QY 1781 ttatcagtgatgtcatctctcaccctcctccaatccttgagaaccccaacccagagcccc 1840  
 DB 483 ttatcagtgatgtcatctctcaccctcctccaatccttgagaaccccaacccagagcccc 542  
 QY 1841 ctgtcccccagctcagccgagatctcaggggtgagcccccaactgtccggaccac 1900  
 DB 543 ctgtcccccagctcagccgagatctcaggggtgagcccccaactgtccggaccac 602  
 QY 1901 gccagccggaccagggaccctccagcgtcaggtgaggtatccccagatgcccgccc 1960  
 DB 603 gccagccggaccagggaccctccagcgtcaggtgaggtatccccagatgcccgccc 662  
 QY 1961 cggccatccaggaacccctgggtctccagctcgtctacttgaaacagtgaggacact 2020  
 DB 663 cggccatccaggaacccctgggtctccagctcgtctacttgaaacagtgaggacact 722  
 QY 2021 ctgcagctcagcgtcgtcgggaaagggtgtcgtggcgccccatttctcctgcattccc 2080  
 DB 723 ctgcagctcagcgtcgtcgggaaagggtgtcgtggcgccccatttctcctgcattccc 782

QY 2081 gtgagtcgggagagaactggatgaacgcagctgtgccgctggtccagggcccccagcct 2140  
 DB 783 gtgagtcgggagagaactggatgaacgcagctgtgccgctggtccagggcccccagcct 842  
 QY 2141 cccctgaacccctgccacgaccccatgcccctactgagcgtggagctggcagtgagcat 2200  
 DB 843 cccctgaacccctgccacgaccccatgcccctactgagcgtggagctggcagtgagcat 902  
 QY 2201 cctgcagcgcctcctgtgtgcccccgccacccagcagcctgcagtcgagtcgagtcgaggaat 2260  
 DB 903 cctgcagcgcctcctgtgtgcccccgccacccagcagcctgcagtcgagtcgagtcgaggaat 962  
 QY 2261 ttggggggggtgctcctcctggtgcccccgagcgtgtggacaattccccggcccaaca 2320  
 DB 963 ttggggggggtgctcctcctggtgcccccgagcgtgtggacaattccccggcccaaca 1022  
 QY 2321 tcaccagctcttgcagcgtgcctctgtggccattgggaagtgtgctctcttgagcc 2380  
 DB 1023 tcaccagctcttgcagcgtgcctctgtggccattgggaagtgtgctctcttgagcc 1082  
 QY 2381 agtgcctcctgctggtgccccgggcccagagagcgcgaggttcgctgtgttggaaca 2440  
 DB 1083 agtgcctcctgctggtgccccgggcccagagagcgcgaggttcgctgtgttggaaca 1142  
 QY 2441 acgggtgatgaagtgaagcagcagagtgctcagggccccccacagccccccagcagag 2500  
 DB 1143 acgggtgatgaagtgaagcagcagagtgctcagggccccccacagccccccagcagag 1202  
 QY 2501 aggcctgtgacatggggccctgtactactgctgttccacagcagtcgagctccaag 2559  
 DB 1203 aggcctgtgacatggggccctgtactactgctgttccacagcagtcgagctccaag 1261

RESULT 2  
 AAD16773  
 ID AAD16773 standard; cDNA; 1020 BP.  
 XX  
 AC AAD16773;  
 XX  
 DT 22-NOV-2001 (first entry)  
 XX  
 DE Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:34.  
 XX  
 KW Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
 KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
 KW autoimmune disease; rheumatoid arthritis; pulmonary disorder; ischaemia;  
 KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
 KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;  
 KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
 KW anti-sense therapy; endocrine disorder; leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Location/Qualifiers  
 FT 65..847  
 FT /\*tag= a  
 FT /product= "Human novel protein"  
 FT /trans\_except= (pos:767..769, aa: Xaa)  
 FT /note= "CDS does not include start codon; Xaa is an  
 FT unknown amino acid"  
 FT /partial  
 XX  
 PN WO200155202-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01325.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.



PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 20-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0215647.  
PR 07-JUL-2000; 2000US-0216680.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-451925/48.  
DR P-PSDB; AA09713.  
XX  
PT Isolated polypeptide for treating, preventing and/or prognosing  
PT medical disorders and also for testing and detection e.g. diagnosis and  
PT screening for agonists -  
XX  
PS Claim 1; SEQ ID No 34; 469pp; English.  
XX  
CC AAD16750-AAD16775 represent cDNAs corresponding to novel human protein

CC genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777  
CC -AAD16780 represent novel human genomic DNA fragments. The novel proteins  
CC and their DNAs are useful for diagnosing, treating, preventing and/or  
CC prognosing inflammatory disorders (bursitis or tendonitis); neural  
CC disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system  
CC disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);  
CC muscular disorders; reproductive disorders; gastrointestinal disorders  
CC (malabsorption syndrome, Crohn's disease); pulmonary disorders;  
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);  
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous  
CC disease and conditions (breast cancer); hyperproliferative disorders  
CC (leukaemia, hyperplasia); tumours; foetal and developmental  
CC abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,  
CC asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine  
CC disorders; pregnancy-related disorders and infections. The novel protein  
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can

Query Match 17.9%; Score 472.4; DB 22; Length 1020;  
Best Local Similarity 99.6%; Pred No. 1.3e-101;  
Matches 484; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2074 atctccctgagtcggagagaaactggatgaacgagctgtgcccgggtgccaggccc 2133  
Db |||||  
QY 33 atctccctgagtcggagagaaactggatgaacgagctgtgcccgggtgccagg-cc 91  
Db |||||  
QY 2134 ccagctccctggaacctgcacgagcaccatgcccctactggaggctggcag 2193  
Db |||||  
QY 92 ccagctccctggaacctgcacgagcaccatgcccctactggaggctggcag 151  
Db |||||  
QY 2194 tggacatctcagcgcctctgtgcccgcgcacccagcaccgcagctgcagtgccg 2253  
Db |||||  
QY 152 tggacatctcagcgcctctgtgcccgcgcacccagcaccgcagctgcagtgccg 211  
Db |||||  
QY 2254 caggaaattgggggggtgctcctcgtgccccggagcgtgtggaatctccccgg 2313  
Db |||||  
QY 212 caggaaattgggggggtgctcctcgtgccccggagcgtgtggaatctccccgg 271  
Db |||||  
QY 2314 cccaatcacccagctctgcccagctgcctctgtgcccattgggaattgctctct 2373  
Db |||||  
QY 272 cccaatcacccagctctgcccagctgcctctgtgcccattgggaattgctctct 331  
Db |||||  
QY 2374 tggagcagctctcctgctgctgctgctgctgctgctgctgctgctgctgct 2433  
Db |||||  
QY 332 tggagcagctctcctgctgctgctgctgctgctgctgctgctgctgctgct 391  
Db |||||  
QY 2434 gggaaacacgggtgatgaagtggagcagagagtgctgctgctgctgctgctgct 2493  
Db |||||  
QY 392 gggaaacacgggtgatgaagtggagcagagagtgctgctgctgctgctgctgct 451  
Db |||||  
QY 2494 agcagagagcctgtgacatggggccctgtactactgttccacagcagctggagc 2553  
Db |||||  
QY 452 agcagagagcctgtgacatggggccctgtactactgttccacagcagctggagc 511  
Db |||||  
QY 2554 tccaag 2559  
Db |||||  
QY 512 tccaag 517

RESULT 3  
AAD16758  
ID AAD16758 standard; cdna; 1998 BP.

AC AAD16758;

XX 22-NOV-2001 (first entry)

XX Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:19.  
DE Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
KW autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;  
KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;

KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
KW anti-sense therapy; endocrine disorder; leukaemia; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
CDS 56..1174  
/\*tag= a  
/product= "Human novel protein"  
/note= "CDS does not include start codon"  
/partial  
XX WO200155202-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01325.

XX 31-JAN-2000; 2000US-0179055.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 06-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 12-SEP-2000; 2000US-0232081.

PR 14-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 21-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-02556719.

PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-451925/48.  
XX P-PSDB; AAE09698.  
DR Isolated polypeptide for treating, preventing and/or prognosing  
DR medical disorders and also for testing and detection e.g. diagnosis and  
DR screening for agonists -  
PT  
PT  
XX Claim 1; SEQ ID No 19; 469pp; English.  
XX  
XX ADP16750-ADP16775 represent cDNAs corresponding to novel human protein  
CC genes, and AAE09690-AAE09715 represent the proteins they encode. AAE16777  
CC -ADP16780 represent novel human genomic DNA fragments. The novel proteins  
CC and their DNAs are useful for diagnosing, treating, preventing and/or  
CC prognosing inflammatory disorders (bursitis or tendonitis); neural  
CC disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system  
CC disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);  
CC muscular disorders; reproductive disorders; gastrointestinal disorders  
CC (malabsorption syndrome, Crohn's disease); pulmonary disorders;  
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);  
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous  
CC disease and conditions (breast cancer); hyperproliferative disorders  
CC (leukaemia, hyperplasia); tumours; foetal and developmental  
CC abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,  
CC asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine  
CC disorders; pregnancy-related disorders and infections. The novel protein  
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can  
CC also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before

Query Match 17.9%; Score 472.4; DB 22; Length 1998;  
Best Local Similarity 99.6%; Pred. No. 1.6e-101;  
Matches 484; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 2074 atctccgtgagtcgggagaggaactggtgaacgacgactgtgcgcgggtgcaggccc 2133  
DB 24 atctccgtgagtcgggagaggaactggtgaacgacgactgtgcgcgggtgcagg-cc 82  
QY 2134 ccagctccctgaacctgccacggccacccatgcccccatactgggaggtggcgag 2193  
DB 83 ccagctccctgaacctgccacggccacccatgcccccatactgggaggtggcgag 142  
QY 2194 tggacatctgcagcgcgtctctgtggcccgccacccacccagccagctgcagtgccgg 2253  
DB 143 tggacatctgcagcgcgtctctgtggcccgccacccacccagccagctgcagtgccgg 202  
QY 2254 caggaatttgggggggtggtctctcgtgcgcggcccgagcgtctgaacatctcccccgg 2313  
DB 203 caggaatttgggggggtggtctctcgtgcgcggcccgagcgtctgaacatctcccccgg 262  
QY 2314 cccaacatcacccagctcttgcagctgcgcctctgtggccattgggaagttggtctctct 2373  
DB 263 cccaacatcacccagctcttgcagctgcgcctctgtggccattgggaagttggtctctct 322  
QY 2374 tggagccagtgctccgtgcggtggccggggccagagagaacccggcaggttcgctgtgtt 2433  
DB 323 tggagccagtgctccgtgcggtggccggggccagagagaacccggcaggttcgctgtgtt 382  
QY 2434 gggaacaacggtgatgaagtgcagcagcagagagtggtgcgtcagggcccccacagccccc 2493  
DB 383 gggaacaacggtgatgaagtgcagcagcagagagtggtgcgtcagggcccccacagccccc 442

QY 2494 agcagagggcctgtgacatggggccctgtactactgctgctgttccacagcagctggagc 2553  
 Db |||||||  
 QY 443 agcagagggcctgtgacatggggccctgtactactgctgctgttccacagcagctggagc 502  
 Db |||||||

QY 2554 tccaag 2559  
 Db 503 tccaag 508

RESULT 4  
 AAH16636  
 ID AAH16636 standard; cDNA; 2964 BP.  
 AC AAH16636;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:15752.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 15752; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SQL Sequence 2964 BP; 705 A; 862 C; 796 G; 601 T; 0 other;

Query Match 6.3%; Score 165.2; DB 22; Length 2964;  
 Best Local Similarity 56.1%; Pred. NO. 4e-29;  
 Matches 333; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 1960 ccgcccatccaggacacccctgggtctccagctgctgactggaaacagtgaggacac 2019  
 Db |||||||  
 QY 2020 tctgctgctcagcgtctctgggaaagggtgtctggcgccccatcttctctctgctctcc 2079.  
 Db |||||||  
 QY 534 acagatgttccagacctgtgggaaaggatcgagctacccctatcttccgctgtgacac 593  
 Db |||||||  
 QY 2080 cgtgagtcgggagaggaactggatgaacgcagctgtgctggcggtggtccagcccccagcc 2139  
 Db |||||||  
 QY 594 agaagcactcatgaagggtctctgagagttactgctccagcatgaagcagaccccc 653  
 QY 2140 tccctgaaacctgccaacgacacccccatgccccccatgctggaggctggcgagtgagaca 2199  
 Db |||||||  
 QY 654 gaggagggagcctgcaacatcttcccttggccagcttctggacatcggggagtggtct 713  
 QY 2200 tctgcaagcgtctctgtggcccgccacacagcagccagctgagtgccgagcagaa 2259  
 Db |||||||  
 QY 714 gagtgcagcaagacctgtggcctgggcatgcagcagccaggtctctgtgcccagggtg 773  
 QY 2260 ttgggggggtggtgctctcctgggtgcccccgagcgtggtggacatctccccggcccaac 2319  
 Db |||||||  
 QY 774 tacgcaacgcagctgacggtgacgacctacgcctgcccagcaccctggagaaacctgag 833  
 QY 2320 ataccagctctgccaagctgcgcctctgtggccattgggaagtgtgctctcttgagc 2379  
 Db |||||||  
 QY 834 accaccagcactgccaactcaagatctgcagcagtggtggcagatccggaccactggacc 893  
 QY 2380 cagtgctccgtgctggtggcgccgagagagcagcagcaggttctgctgtgtgggaac 2439  
 Db |||||||  
 QY 894 tcgtgctggtgctgctggcggtgggacagagaccctggtgagagtggtgagcaac 953  
 QY 2440 aacggtgatgaagtgagcagcagagtggtgctcagggccccccacagccccccagaga 2499  
 Db |||||||  
 QY 954 attgggatgtgtgacgatgaggaatgcaacatgaagctc---cgccgaatgacatt 1010  
 QY 2500 gaggcctgtgacatggggccctgactactgctgctggttccacagcagctggagc 2553  
 Db |||||||  
 QY 1011 gagactgcagatgggaccctgtgccaagagctggttctccaccgagtgagc 1064

RESULT 5  
 AAH99398  
 ID AAH99398 standard; cDNA; 704 BP.  
 AC AAH99398;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein encoding cDNA sequence SEQ ID NO:233.  
 XX  
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiagregant; haemostatic; vulnery; antileuc; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder; ss.

XX Homo sapiens.  
 XX WO200153455-A2.  
 XX PD 26-JUL-2001.  
 XX PF 22-DEC-2000; 2000WO-US35017.  
 XX PR 23-DEC-1999; 99US-0471275.  
 XX PR 21-JAN-2000; 2000US-0488725.  
 XX PR 25-APR-2000; 2000US-0552317.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-457603/49.  
 XX DR P-PSDB; AAM25457.  
 XX Isolated human polynucleotides encoding polypeptides, useful for the  
 XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.  
 XX Claim 1; Page 401-402; 1217pp; English.  
 XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 XX AAM25963. The proteins can have activities based on the tissues and  
 XX cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
 XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 XX cardiovascular; antianemic; antiaggregant; haemostatic; vulnerary;  
 XX antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
 XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 XX encoding them can be used in gene therapy, antisense therapy and vaccine  
 XX production. The proteins and polynucleotides are useful for screening for  
 XX agonists or antagonists of a protein and for the treatment and diagnosis  
 XX of disorders associated with the activity of a protein e.g. inflammation,  
 XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 XX infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 XX osteoporosis, severe combined immunodeficiency, eczema, allergic  
 XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 XX Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 XX neurological disorders.  
 XX Sequence 704 BP; 184 A; 158 C; 195 G; 167 T; 0 other;

Query Match 4.5%; Score 118.2; DB 22; Length 704;  
 Best Local Similarity 58.0%; Pred. No. 3.1e-18;  
 Matches 229; Conservative 0; Mismatches 163; Indels 3; Gaps 1;

QY 1371 gagcccgctgtgagatccttggctggtcggagcgtcctgagctgtggtgagctgtg 1430  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||  
 Db 216 gagcattggtgtgactactaggtccgacaaagtctggtgacaaatgtgggtgtg 275  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 QY 1431 tgggggtgattacattcgtcgtctgttctgggggaacctcaactaccgagggggccc 1490  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 276 tggaggagacaacacgggtgtcaggtgtgtcggggtgtttaaagcatgcctccaccag 335  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 QY 1491 cctgggctacagaagatttggattccagcggagccttgcggctccagattgccca 1550  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 336 cctgggctacacccggtgtgagattcccgaggagccagcaaaataacatacagga 395  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 QY 1551 gtcctggcctagctcaactcctggtgacattcgttgccttggggccggtccatcataa 1610  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 396 gatgtacaagagacaacaatttggccctgagaagcttcttgacgtccatcatacaa 455  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 QY 1611 tggggaactgggtgtggatccccctgggtccctacagggccggtggagacgtcttcgata 1670  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 456 tggggaactgggcaattgatcgcaccaggaataacagagggcgaggaccatgttcaccta 515  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1671 taaccgtctccagagagagagagca--aagggagagctctgtcgcgtgaagggccccc 1727  
 || ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||  
 Db 516 caagcgtccaaatgagatttcgcgcactgcgcgagagctctcttttggcgaaggtccccc 575  
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1728 caccagcctgtgagctgtctatgatcttccagg 1762  
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 576 caacgagatcttgatgtctacgtgatttgatg 610  
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6  
 AAS97176  
 ID AAS97176 standard; cDNA; 2805 BP.  
 XX AAS97176;  
 AC AAS97176;  
 XX 26-FEB-2002 (first entry)  
 XX Human metalloprotease partial DNA sequence #5.  
 XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;  
 KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;  
 KW hypertensive; antispasmodic; neuroleptic; neuroprotective; anabolic;  
 KW anorectic; antiinflammatory; aspartyl protease; cysteine protease; colon;  
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; lung;  
 KW lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;  
 KW immune-related disease; cardiovascular disease; neuronal disease;  
 KW migraine; sexual dysfunction; mood disorder; attention disorder;  
 KW cognition disorder; hypotension; hypertension; psychotic disorder;  
 KW dyskinesia; metabolic disorder; inflammatory disorder; ss.  
 XX Homo sapiens.  
 OS WO200183782-A2.  
 XX 08-NOV-2001.  
 XX 04-MAY-2001; 2001WO-US14431.  
 XX 04-MAY-2000; 2000US-201879P.  
 XX (SUGEN-) SUGEN INC.  
 XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepel S;  
 PI Payne V;  
 XX WPI; 2002-041502/05.  
 XX P-PSDB; AAU72893.  
 XX Novel protease polypeptide useful for screening for substances that may  
 XX be used to treat, e.g., cancers, immune-related diseases,  
 XX cardiovascular disease, migraine, pain, psychotic and inflammatory  
 XX disorders -  
 XX Claim 30; Figure 11-J; 232pp; English.  
 XX The invention relates to an isolated, enriched, or purified protease  
 XX polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to  
 XX screen for substances (S) that may modulate its activity. Administering  
 XX S (which modulates protease activity in vitro) may be used to treat a  
 XX disease or disorder selected from cancers (e.g., of tissues, of blood or  
 XX haematopoietic origin, of the breast, colon, lung, prostate, cervical,  
 XX brain, ovarian, bladder or kidney), immune-related diseases and  
 XX disorders, cardiovascular disease, brain or neuronal-associated diseases  
 XX (e.g., central or peripheral nervous system diseases, migraine, pain,  
 XX sexual dysfunction, mood disorders, attention disorders, cognition  
 XX disorders, hypotension, hypertension, psychotic disorders, neurological  
 XX disorders and dyskinesias), metabolic disorders and inflammatory  
 XX disorders. (I) may also be useful as a diagnostic tool for a disease or  
 XX disorder such as those above. AAS97159-AAS97195 represent human  
 XX protease coding sequences and primers of the invention.  
 XX Sequence 2805 BP; 745 A; 577 C; 733 G; 650 T; 0 other;

XX WO200183782-A2.  
PN XX  
PD XX  
PF 08-NOV-2001.  
XX  
XX 04-MAY-2001; 2001WO-US14431.  
XX  
XX 04-MAY-2000; 2000US-201879P.  
PR XX  
XX (SUGEN-) SUGEN INC.  
PA XX  
XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
PI Payne V;  
PI  
XX WPI; 2002-041502/05.  
DR P-PSDB; AAU72891.  
DR  
XX Novel protease polypeptide useful for screening for substances that may  
PT be used to treat, e.g., cancers, immune-related diseases,  
PT cardiovascular disease, migraine, pain, psychotic and inflammatory  
PT disorders.  
XX  
XX Claim 30; Figure 1H; 232pp; English.  
PS  
XX The invention relates to an isolated, enriched, or purified protease  
CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to  
CC screen for substances (S) that may modulate its activity. Administering  
CC S (which modulates protease activity in vitro) may be used to treat a  
CC disease or disorder selected from cancers (e.g., of tissues, of blood or  
CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,  
CC brain, ovarian, bladder or kidney), immune-related diseases and  
CC disorders, cardiovascular disease, brain or neuronal-associated diseases  
CC (e.g., central or peripheral nervous system diseases, migraine, pain,  
CC sexual dysfunction, mood disorders, attention disorders, cognitive  
CC disorders, hypotension; hypertension, psychotic disorders, neurological  
CC disorders and dyskinesias), metabolic disorders and inflammatory  
CC disorders. (I) may also be useful as a diagnostic tool for a disease or  
CC disorder such as those above. AAS97159-AAS97195 represent human  
CC protease coding sequences and primers of the invention.  
XX  
SQ Sequence 3675 BP; 874 A; 1009 C; 1019 G; 773 T; 0 other;

Query Match 4.3%; Score 113.2; DB 24; Length 2805;  
Best Local Similarity 50.0%; Pred. No. 6.9e-17;  
Matches 317; Conservative 0; Mismatches 308; Indels 9; Gaps 1;  
QY 1159 cgggcccgcagtcgagcccttaactccaggaattcattcagccagctgtatcagtg 1218  
DB 1039 cgggctcaacagtgacagataacagcaaaccttccgtggatgtttaccagtg 1098  
QY 1219 gagccttcactgaagtcagcgtccagcgtgtgacgaactcgcggcccgctggc 1278  
DB 1099 aaacctatacaaaagtggaggaagatcgatcaaacactgactcaagctgagac 1158  
QY 1279 ttccgctttatgtccgtccactgaaaggccagggatccagccctgtgacgctgga 1338  
DB 1159 ttggaatttttttgcaatgtccggcaaaagtgaagatgaactccctgtcccaaac 1218  
QY 1339 gccctgacatctgtgtgctgagcgtgtgagcccggtgtgagtgagctctggc 1398  
DB 1219 aaaaatgatgtttgatgacgggtttgtgaactagtggaatgcatgaactagc 1278  
QY 1399 tctgcagggcgtccctgatggctgtgagtcgtgtgggtgtgattctacctgtgcctt 1458  
DB 1279 tctaaagcagtttccagatctgtgtgctgttgcaaaagtgataatcaactgcaagtt 1338  
QY 1459 gtttcggggaacctcactgacgcagggggccctcctgggtatcagaagatctgtggatt 1518  
DB 1339 tataaagcctgtactcaaccagcataaagcaaatgaatattacccgtgtgctcatt 1398  
QY 1519 ccagcgggagccttcggtccagattgccagctcccgctcagtcacactacactggca 1578  
DB 1399 ccagctggccgcagcatcgaaatccaggagctgcaggtttctccagttacactcgca 1458  
QY 1579 ctctggtccctggggcggtccatcatcaatggaaactgggtgtgattccctcctggg 1638  
DB 1459 gttcgaagcctcagtcocaaagtattaccctaccggggcgtggagcagctggtcctgg 1518  
QY 1639 tectcagggcggggggaccgtcttttgatataaccctcccccagggagggagcaca 1698  
DB 1519 gattcccccttcgctgggaccacgtttgaataccagcgtctttcaa-----ccgc 1569  
QY 1699 ggggagatctctcggtgaagcccccacacccagcctgtgagtgctatatgattctt 1758  
DB 1570 ccggacgtctgtacgcgcagggcccaacaaatgagacgctggtttgaaattctgatg 1629  
QY 1759 caggaggaacacccagcgcttttttatcagtatg 1792  
DB 1630 caaggcaaaatccaggagtagcttggaagtatg 1663  
RESULT 7  
AAS97174  
ID AAS97174 standard; CDNA; 3675 BP.  
AC AAS97174;  
XX  
XX 26-FEB-2002 (first entry)  
DT  
XX Human metalloprotease partial DNA sequence #3.  
DE  
XX Human; protease; PCR primer; cytosolic; immunomodulator; cardiac;  
KW vasotrophic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;  
KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;  
KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;  
KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;  
KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;  
KW immune-related disease; cardiovascular disease; neuronal disease;  
KW migraine; sexual dysfunction; mood disorder; attention disorder;  
KW cognition disorder; hypotension; hypertension; psychotic disorder;  
KW dyskinesia; metabolic disorder; inflammatory disorder; ss.  
XX Homo sapiens.  
OS







Best Local Similarity 47.6%; Pred. No. 8.1e-13;  
Matches 323; Conservative 0; Mismatches 350; Indels 6; Gaps 1

QY	1118	cttgagccaagcgccctgccccctgagcagccagaccgcccgccctgcagtcgcgcag	1177
Db	8870	ctgcacaaccacagctgccgaggaagaatccgactccggcgagcagtgctccc	8929
QY	1178	ctttaactcccaagaaattcatgggcoagctgtatcagtgaggagccctcaactgaagtc	1237
Db	8930	gtcttcacaggcagcagttgatggattctctcagatgggtgcgtacacgaatgcac	8989
QY	1238	agggctccacgcgtgtaactgaactcggccgcctggtctccgcttctatgtccgtc	1297
Db	8990	-----cgaatccctcgagctgaactgtatgccaaggcgagggttctactaccgtc	9043
QY	1298	acactgaaagctccaggatgggacctgtgtcagctgtgagccctgacactctgtgtg	1357
Db	9044	agcgggagaagtgtgtgacgaaccgcgtcaacgacaaggacctgcagtgctgta	9103
QY	1358	ctgacgctgtctgagccccggctgtgatgggacctgtgctctgagcagcgctctg	1417
Db	9104	acggggatgcatgcgggtgggtgcgacatgctgggtagcgtgccaaaggagaca	9163
QY	1418	gctgtgagctgtgggggtgatgattctactctgcctgtttctggggaaacctcactg	1477
Db	9164	agtgcgaagtgtgtgtgtgtgtggcagcaactcgaagacattcgcaacacatcacca	9223
QY	1478	accgagggggccccctgggctatcagaagatctgttgattccagcggggagcctgcggc	1537
Db	9224	cgaagatctgctctgtgatacaatgactgtgtcttccagagggagccactaaca	9283
QY	1538	tcagattgccagctccggcctagctccaactacctggcaactctgtgccccctggggcc	1597
Db	9284	ttcgtatcaggagactgtgccctccagcaatattacctggcgtccggaaccacagtcgac	9343
QY	1598	ggtccatcatcaatgggaactggctgtggatccccctggctcctacagggccggcgga	1657
Db	9344	actactacctgaacggcatgtggcgcacgcacttcccaacgcccatgtcttcgcgaact	9403
QY	1658	ccgtctcttcgatatcaactcctccagggagggcggaagggagagtcctctcggctg	1717
Db	9404	ctgtgtgggaactaccaacgaaagccaatgggtcttcgcgcctcccgacaactgacctgca	9463
QY	1718	aaggccccaccaccagcctgtggatgtctctatgatcttccaggaggaagaaacccaggcg	1777
Db	9464	gcggtcccatctcgagagcctcttcattgattatgttggtcgaggaagaacatcagcc	9523
QY	1778	tttttatcagtatgtcat	1796
Db	9524	tcgaactacgatacagcat	9542

RESULT 11  
AAF82149  
ID AAF82149 standard; cDNA; 5061 BP.

RESULT	11	
AAF82149		
ID	AAF82149	standard; cDNA; 5061 BP.
XX	XX	
XX	AAF82149;	
XX	XX	
DT	02-JUL-2001	(first entry)
XX	XX	
DE	Human ADAM	type metal protease MDTs1 encoding cDNA SEQ ID NO:2.
XX	XX	
KW	Human;	a disintegrin and metalloprotease type metal protease; MDTs1;
KW	MDTs2; MDTs3;	ADAM type metal protease; cytosstatic; antiarthritic;
KW	cancer; arthritis;	arthrosis deformans; ss.
XX	XX	
OS	Homo sapiens.	
XX	XX	
FH	Key	Location/Qualifiers
CD	1..5061	
CD	/*tag- a	
FT	/product= "MDTs1"	
FT	/note= "a disintegrin and metalloprotease (ADAM)	
FT	FT	

```

type metal protease"
JP2001008687-A.
16-JAN-2001.
25-JUN-1999; 99JP-0180973.
25-JUN-1999; 99JP-0180973.
(YAMA ) YAMANOUCHI PHARM CO LTD.
WPI: 2001-285362/30.
P-PSDB; AAB82149.
New metal protease and metal protease gene, for use as a drug for
cancers, arthritis and arthrosis deformans -
Example 1; Page 17-19; 31pp; Japanese.
The present sequence encodes a disintegrin and metalloprotease (ADAM)
type metal protease designated MDRS1, isolated from human. MDRS proteins
have cytostatic and antiarthritic activities. They can be used as a drug
for cancers, arthritis and arthrosis deformans.
Sequence 5061 BP; 878 A; 1746 C; 1587 G; 850 T; 0 other;

Query Match          3.7%; Score 96.6; DB 22; Length 5061;
Best Local Similarity 51.3%; Pred. No. 6.6e-13;
Matches 368; Conservative 0; Mismatches 319; Indels 30; Gaps

```

Query Match	3.78;	Score 96.6;	DB 22;	Length 5061;
Best Local Similarity	51.3;	Pred. No. 6.6e-13;		
Matches 368;	Conservative 0;	Mismatches 319;	Indels 30;	Gaps 5;
Qy 1090	tgttctggggagagtgaacagctgaagacgtcgacgcaagccgctgcccccctgagcag	1149		
Db 1729	tgttgggtgagcgcaagcggttcgcgtctgaacctgagcgtgcctgtgctgctgcac	1788		
Qy 1150	ccagaccggggccctgcagtgcgacgcttttaactcccaggaattcatgggcacgtg	1209		
Db 1789	ccctcttccgcgaagtccagtgcagccattgacgctatgctctacaaggccagctg	1848		
Qy 1210	tatacgttgggagcccttaactgaagtcacgggctcccgagcgtctgaaactgaactgcgg	1269		
Db 1849	cacacatgggtgcccgtg-----gtcaatgacgtgaacccctgcgactgcactgcgg	1902		
Qy 1270	ccccgtggctccgctttctatgtccgtcacactgaaaggctccaggatgggaccctgtgt	1329		
Db 1903	cccgcaatgagtactttgcgagaaagtgcggagcgcgtgggtcgatggcacccctgc	1962		
Qy 1330	cagcctggagc-----ccctgaacatctgtgtggctggacgctgtctgagcccggtctg	1383		
Db 1963	taccaggtccagaccgacggggaacctgcgcatcaacggcactctgtaagaaactgggctg	2022		
Qy 1384	gatggatccttggtctctggcagcgctcctgatggctgtgagctctgtgggggtgatgat	1443		
Db 2023	gacttcgagattgactccggtgctatggaggacgcgtgtgtgtgtgcacggcgaacgc	2082		
Qy 1444	tctacctgccttgtttcgggaaacctcaactacgagggggcccccctgggctatcag	1503		
Db 2083	tcacctgtccacacggtagcggaaccttcagagggccgaggg-----cctgggggtatgtg	2139		
Qy 1504	aagatcttgtgattccagcggagccttgcggtccagattgcccagctccggcctagc	1563		
Db 2140	gatgtggggctgatccacgcggcgacgagatccgcattccaaagaggttgcggaggt	2199		
Qy 1564	tccaactaccggcactc-----gtggccctggggccggtccatcatcaatgggaactgg	1620		
Db 2200	gccaaactcccggaactgcggagtgaggacccggagaaagtacttctcaatgtgtgctgg	2259		
Qy 1621	gctgtggatcccccttggttcctacagggccggcgaggcgtcttttcgataataacgtct	1680		
Db 2260	accatccagtgaacgggagatccaggtgtagcgggacaccttcaacatgcgcag-----	2315		
Qy 1681	cccagggagggggcaagagggagagctgtctcggtctgaagcccccacacccagcctgtg	1740		

```

Db 2316 -----caggagcaactggagaaactccagctcccggtcccaacaggagcctgtc 2367
Qy 1741 gatgtctatatttccagaggaacccagcgcttttttttcaatgatgtatc 1797
Db 2368 tggatccagctgttccagagagcaacccctgggtgcatacagatcacaccatc 2424

RESULT 12
AD004475
ID AAD04475 standard; cDNA: 5353 BP.
XX
AC AAD04475;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human 27875 ADAM-TS cDNA, alternative version.
XX
KW Human; ADAM-TS, A Disintegrin And Metalloproteinase; antiinflammatory;
KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW cardiant; tumour; thymoma; vasotropic; cytostatic; virucide; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 36..5096
FT /tag= a
FT /product= "Human 27875 ADAM-TS protein"
FT /transl_except= (pos:1821..1823, aa:Xaa)
FT /note= "Xaa is an unknown amino acid"
FT sig_peptide 36..125
FT /tag= b
FT mat_peptide 126..5093
FT /tag= c
FT /product= "Human mature 27875 ADAM-TS protein"
XX
WO200131034-A1.
XX
XX
XX
XX
XX 25-OCT-2000; 2000WO-US29380.
XX
XX 25-OCT-1999; 99US-0426282.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Kapeller-Libermann R, White D;
XX
XX WPI; 2001-300513/31.
XX
XX P-PSDB; AAEC0913.
XX
XX Novel isolated polypeptide, 27875, a human ADAM-TS (a disintegrin and
XX metalloproteinase) useful for diagnosis and treatment of disorders of
XX bone, lung, heart, skeletal muscle such as osteoporosis, emphysema,
XX angina
XX
XX Claim 7; Page 115-123; 133pp; English.
XX
XX The present sequence is an alternative version of a cDNA encoding
XX 27875 protein, a human ADAM-TS (a disintegrin and metalloproteinase).
XX Metalloproteinase is a proteolytic enzyme involved in protein
XX maturation, protein degradation, tumour growth, metastasis and
XX angiogenesis. Nucleotides encoding 27875, 27875 protein and its
XX antibodies are useful for preventing, diagnosing and treating
XX 27875 metalloproteinase-related disorders. These disorders
XX include, disorders involving the lung such as congenital anomalies,

```

```

CC bronchial asthma, Goodpasture's syndrome, pulmonary alveolar proteinosis,
CC disorders involving heart such as heart failure, cardiac hypertrophy,
CC angina pectoris, myocardial infarction, chronic ischaemic heart disease,
CC disorders involving the skeletal muscle include tumours such as
CC rhabdomyosarcoma, disorders involving blood vessel such as hypertension,
CC atherosclerosis, vasculitis associated with other disorders, disorders
CC involving the testis and epididymis such as nonspecific epididymitis and
CC orchitis, gonorrhea, mumps, tuberculosis and syphilis, spermatocytic
CC seminoma, disorders involving the kidney such as membranoproliferative
CC glomerulonephritis, necrotising glomerulonephritis, renal artery
CC stenosis, chronic glomerulonephritis, sickle cell disease nephropathy,
CC urinary tract obstruction, disorders of the bone such as achondroplasia,
CC osteoporosis, Paget's disease, rickets, osteomalacia, osteonecrosis,
CC tuberculous osteomyelitis, osteoma, osteosarcoma, metastatic tumours,
CC and disorders involving thymus such as thymic cysts, thymic hypoplasia,
CC thymomas, and Hodgkin's disease. 27875 sequence is used in gene therapy.
CC Note: This sequence SEQ.ID.NO.2 is stated as being similar to the
CC sequence shown in Fig 1 (See AAD04494). However these sequences differ
CC at several locations.
XX
SQ Sequence 5353 BP; 960 A; 1833 C; 1648 G; 911 T; 1 other;
XX
Query Match 3.7%; Score 96.2; DB 22; Length 5353;
Best Local Similarity 51.2%; Pred. No. 8.4e-13;
Matches 367; Conservative 1; Mismatches 319; Indels 30; Gaps 5;
Qy 1090 tgttctggggagtgaaacagctgaagcctgcagcgaagcgcctcccccctgaagcag 1149
Db 1764 tgtgtgggtgagcgcaagcgttcctcctgcactgcaggcctgcctgtggcgcc 1823
Qy 1150 ccagacccccgggcccctgcagtcgcagccttaactcccagggaattcattgagccagctg 1209
Db 1824 cctctctcccgccagctccagtcagccactttgacgttatgtctctcaaggcgccagctg 1893
Qy 1210 tatcagtgggagccttcactcaatgaagtcaggggtcccgagcgtgtgaactgaactgcgg 1269
Db 1884 cacacatgggtgcccgtg-----gtcaatgcgtgaaccctgcgagctgcactgcgg 1937
Qy 1270 cccggtggtccctctctctatgtccgtcacactgaaaggtccagagtgagaccctatgt 1329
Db 1938 ccgcgaatgagctatttgcggaagcgtgcggaagcgcgtgcgtatggcccccctgc 1997
Qy 1330 cagcctggagc-----ccctgacatctgtgtggctggaacgctgtctgagccccggtgt 1393
Db 1998 taccaggtccgagccagcgcgggaacctctgcatacagcgcattctgtaagaaacgtggctgt 2057
Qy 1384 gatgggaccttgctctggtcagcgctcctctgctgctgagtgagtgctgtgggggtgatgat 1443
Db 2058 gacttcgagattgactccgtgtctatggagagcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2117
Qy 1444 ttactctgcgccttcttcgtgggaacctcactgaccgagggggcccccctgggctatcag 1503
Db 2118 tccacctccacacgtgagcgggacaccttcgagagggccgaggg---cctggggtatgtg 2174
Qy 1504 aagatctgttgattccagcgggagccttcggttcacagattgccagctccgccttagc 1563
Db 2175 gatgtgggctgatcccgccggcgacgcagcagatccgcatcccaagaggttgccgagggct 2234
Qy 1564 tccaaactacctggcacttc---gtggccctggggccggtccatcatcaatgggaactgg 1620
Db 2235 gccaaactcctggcactcggagtgaggagcccgagagagtaacttctcctcaaatgggtgtg 2294
Qy 1621 gctgtggatccccctgggtctctacagggccggcgccggtcttttcgataataacogtct 1680
Db 2295 accatccagtggaacggggagactaccaggtggcagggaaccaacctcaacacgacg---- 2350
Qy 1681 ccaggggagggggaacaaaggagagtgctgcggtgagggcccccacccacccctgtg 1740
Db 2351 -----cagggggcaactggggaacctcactgcctcccgggtcccacacagagagcgtc 2402
Qy 1741 gatgtctatgatctttcagagagaacacccagcgcttttttatcagtatgtcatc 1797

```







Db 1966 gctgaaggttataattctacactgaactgctctcggtgatcgatgggacccagtc 2025  
QY 1330 cagcctggagccctgacatctgtgtggtgacgctgtctgagcccccgtgtgtgatggg 1389  
Db 2026 aatgcggtattcactgatactgcatcaatggagaatgcaagcacgtaggtgtgataat 2085  
QY 1390 atccttggctctgagcggcgtctctgatggctgtggagtctgtgggggtgatattctacc 1449  
Db 2086 attttggatctgatctagggagatagatgtcgagtctgtggagggacggaagcaca 2145  
QY 1450 tgtgccttgttttcggggaaacctcaactgaaccgagggggccctctgggtatcagaagatc 1509  
Db 2146 tgtgatgccattgaagggttcttcaatgattcactgccaggggaggtacatgggaagtg 2205  
QY 1510 ttgtgaattccagcgagagccttgcggctccagattgccagctccggcctagctccaac 1569  
Db 2206 gtgcagataaccaagaggtctgttccattgaagttagagaagttgccatgtcaagaac 2265  
QY 1570 taactggcaactctgtggccctggggccggtccatcataatgggaactgggctgtggat 1629  
Db 2266 tatattgctttaaaatctgaaggagatgattactatattaatgtgtgctggactattgac 2325  
QY 1630 cccctgggtccctacagggccgggacccgtcttttcgatatataacgctcc 1679  
Db 2326 tggcctaggaaatttgatgttctgtggacagcgttttcattacaagagacc 2375

Search completed: July 23, 2002, 17:26:49  
Job time: 5379 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 17:36:00 ; Search time 3257.62 Seconds  
(without alignments)  
16920.481 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atggagaactgactggcag.....cctcagcctttccagcatag 2634

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.un.\*
- 28: em.vi.\*
- 29: em.htg\_hum.\*
- 30: em.htg\_inv.\*
- 31: em.htg\_other.\*
- 32: em.htg\_inv.\*
- 33: em.htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

length of match

Result No. Score Match length DB ID Description

1	2508	95.2	4230	6	AX342636	AX342636 Sequence
2	699	26.5	176550	9	AL356356	AL356356 Human DNA
3	529	20.1	2070	9	AF217974	AF217974 Homo sapi
4	457	17.4	207815	2	AC053497	AC053497 Homo sapi
5	205	7.8	207815	2	AC053497	AC053497 Homo sapi
6	39	1.5	142023	2	AC092479	AC092479 Mus muscu
7	39	1.5	215810	2	AC093317	AC093317 Mus muscu
8	30	1.1	1409	10	BC016215	BC016215 Mus muscu
9	23	0.9	2703	9	BC008840	BC008840 Homo sapi
10	23	0.9	2964	9	AK023772	AK023772 Homo sapi
11	23	0.9	62385	2	AC101188	AC101188 Mus muscu
12	23	0.9	151333	2	AC026636	AC026636 Homo sapi
13	23	0.9	184483	2	AC104938	AC104938 Homo sapi
14	22	0.8	168872	2	AC019036	AC019036 Homo sapi
15	22	0.8	175970	9	AC097488	AC097488 Homo sapi
16	21	0.8	4506	2	AC103667	AC103667 Homo sapi
17	21	0.8	45257	2	AC084177	AC084177 Homo sapi
18	21	0.8	67047	2	AC013544	AC013544 Homo sapi
19	21	0.8	72907	2	AC016010	AC016010 Homo sapi
20	21	0.8	138706	9	AF003469	AF003469 Homo sapi
21	21	0.8	148427	2	AC018981	AC018981 Homo sapi
22	21	0.8	168811	2	AC018941	AC018941 Homo sapi
23	21	0.8	168811	2	AC018941	AC018941 Homo sapi
24	21	0.8	168601	2	AC016803	AC016803 Homo sapi
25	21	0.8	183979	2	AL354671	AL354671 Homo sapi
26	21	0.8	189046	2	AL592301	AL592301 Homo sapi
27	21	0.8	209844	9	AC011495	AC011495 Homo sapi
28	21	0.8	215441	9	AC005736	AC005736 Homo sapi
29	20	0.8	1189	6	AX061802	AX061802 Sequence
30	20	0.8	1326	5	DR043898	DR043898 Danio rerio
31	20	0.8	1575	9	AK026179	AK026179 Homo sapi
32	20	0.8	1927	10	AB040490	AB040490 Mus muscu
33	20	0.8	1979	5	DRU77736	DRU77736 Danio rerio
34	20	0.8	2189	5	AF039952	AF039952 Danio rer
35	20	0.8	2239	9	AK057145	AK057145 Homo sapi
36	20	0.8	2562	9	BC018923	BC018923 Homo sapi
37	20	0.8	2671	9	AB056152	AB056152 Homo sapi
38	20	0.8	2994	4	DR2P12POU	DR2P12POU
39	20	0.8	3160	8	BFBP8GEN	BFBP8GEN
40	20	0.8	3215	10	RNU96920	RNU96920
41	20	0.8	3246	9	AF218313	AF218313 Homo sapi
42	20	0.8	3840	10	MMFBLN2S01	MMFBLN2S01
43	20	0.8	4398	10	RNU96921	RNU96921 Rattus norv
44	20	0.8	38399	2	AL354885	AL354885 Homo sapi
45	20	0.8	47338	9	AL355990	AL355990 Human DNA

ALIGNMENTS

RESULT	1					
AX342636						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
JOURNAL						
FEATURES						
Source						

AX342636 Sequence 33 from Patent WO0198468.  
AX342636 4230 bp DNA linear PAT 12-JAN-2002  
AX342636.1 GI:18152033

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (sites)  
Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,  
Tribouley, C.M., Deleage, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,  
Hafalla, A., Khan, F.A., Walla, N.K., Yao, M.G., Lu, D.A., Patterson, C.,  
Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R.  
Patent: WO 0198468-A 33 27-DEC-2001;  
Incyte Genomics, Inc. (US)  
Location/Qualifiers  
1. 4230  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7604035CB1"

BASE COUNT	810	a	1411	c	1210	g	799	t
ORIGIN								
Query Match	95.2%; Score 2508; DB 6; Length 4230;							
Best Local Similarity	100.0%; Pred. No. 0;							
Matches 2558; Conservative	0; Mismatches 1; Indels 0; Gaps							
QY	1	atggagaactggactggcaggccctggcgctgtatctgcgtgcttctgttcctcctcctcaag	60					
DB	228	ATGGGAANCTGGACTGGCAGGCCCTGGCTGTATCTGCTGCTGCTTCTTCCTCCCTCCCTCAG	287					
QY	61	ctctgcttgatcacgaggtgtgtctcgacacctctctcagacaacctcacagagagggc	120					
DB	288	CTCTGCTTGATCATGAGAGGTGTGTCCGCACACTCTCTTCAGACACCTTACAGAGAGAGGGC	347					
QY	121	caggcccccaaggtgtctctgggaaccttgggtctcagtgggcoctttgtctccagccctgc	180					
DB	348	CAGGGCCCCGAAGGTGCTCTGGGACCTTTGGGTCCAAGTGGGCTCTTGTCTCCAGCCCTGC	407					
QY	181	ggggtgggggtcagcgcagcgcgcgcacatgtcagctccctcacgtccagctcacaccg	240					
DB	408	GGGTTGGGGTGCAGGCCAGAGACC GGACATGTCAAGTCCCCTPACAGTAGTCAGCTCACCCG	467					
QY	241	agttgccctcccctcccggcccccaagacatccagaagccctctcccgcgggcccag	300					
DB	468	AGTCTGCCCTCCCTCCCGGGCCCCAAGACATCCAGAAGCCCTCTCTCCCGGGGGCCAG	527					
QY	301	ggtccagaccccagacttctccagaaacctctcccttgtacaggacacagtctcgggga	360					
DB	528	GGTCCACAGACCCAGAGCTTCTCCAGAAACCCCTCCCTTGTACAGGACACAGTCTCGGGGA	587					
QY	361	aggggtggccactctcagttgcggtctccaccttagggagagagagaccagagatt	420					
DB	588	AGGGTGGCCCACTTCCAGSTCCCGCTTCCCACCTTAGGGAGAGGAGACCCACAGAGATT	647					
QY	421	cgaagcggcaggaggtcccggtctcgagaccocatcaagccaagaatgttcggttatggg	480					
DB	648	CGAGCGGCCAGAGGFTCCCGGCTTCGAGACCCCATTAAGCCAGGAATGTTCCGTATTGGG	707					
QY	481	agsgtgcctttgcattgcactgcacgcgaaccgcagccctcggagcccaccaga	540					
DB	708	AGAGTGCCCTTTGCATTGCCATGTCAACGGGAACCGACGCCCTCTGGAGGCCACCCACA	767					
QY	541	tctgagctgcctgatctctctctaggggggaagggtatatccogtcccctactccaaga	600					
DB	768	TCTGAGCTGTCCCTTGATCTCTTCTAGAGGGGAAGAGCCTATTCCGCTCCCTTACTCCAAGA	827					
QY	601	gcagagccattctcgcgaacggcagcccaaacctgagctccctcccaagaactctct	660					
DB	828	CGAGACCCATTCTCGCAAACGGCAGCCCCAACTGAGCTGCCCTTCCCACAGAACTGTCT	887					
QY	661	gtccacaccccatcccccaagcgaaccttaagccctgaacctgaacagagaggtg	720					
DB	888	GTCCACACCCCATCCGCCCAAGCAGAACTCTTAAGCCCTGAAACTGCTCTAGACAGAGGTG	947					
QY	721	gccccagaaacaggctctccctcctacgcatcacccagagcccaggcoctctgggaca	780					
DB	948	GCCCCAGAACCAAGCTGCCCCCTACGGCATCACCCAGAGGCCACAGGCGCTCTGGCACA	1007					
QY	781	gagccccctcaccacgcactctcttaggaaggtgcttcttcgttcgatccccctcag	840					
DB	1008	GAGCCCCCTCACCCACGCACTCTCTTAGAGAAAGTGTGCTTCTTCCGTGCATCCCTCTCAG	1067					
QY	841	ccacgaagcceaagttcccaggttggccagttcccaggttagcaggaggaacgcctcgt	900					
DB	1068	CCACGAAGGCCAAGTTCOCAGGTTGGGCCAGTCCCCAGSTAGCAGGGGACAGCGCCCTGAT	1127					
QY	901	cttttctctcgttccctcgggcccgcagcagcaggccaaggcccttggggaacgggg	960					
DB	1128	CCTTTTCTTCGTTCCGTCGGGGCCGAGGCCACAGAGGCCCAAGGGCCTTGGGNAACGGGG	1187					
QY	961	gggaactctcaaggcccccgcctggagcctgacctcagcaccggcgccgtcgtgctgcc	1020					



```

QY 1033 gccccatgccagctccctctgagcctcttctgctccagtagccctattccaaagtgt 1092
Db 81286 GCCCCCAATGACACTCCCTCTGAGAGCTCTTGTCTCCAGTAGCCCTATTCAGATGT 81345
QY 1093 tctgggagagtgaaacagctaaagcctgcagccaagcg 1131
Db 81346 TCTGGGAGAGTGAACAGCTAAGAGCCTGCAGCCAAGCG 81384

RESULT 3
AF217974 Homo sapiens 2070 bp mRNA linear PRI 01-OCT-2000
LOCUS AF217974 Homo sapiens clone P11396 unknown mRNA.
DEFINITION AF217974
ACCESSION AF217974
VERSION AF217974.1 GI:10441878
KEYWORDS FLI CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
AUTHORS Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu.Y.,
Yu,J. and Han,L.H.
TITLE Novel Human CDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2070)
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu.Y.,
Yu,J. and Han,L.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related
genes, Shanghai Cancer Institute, 25 Ln 2400, Xie-Tu Road, Shanghai
200032, People's Republic of China
FEATURES
source
1..2070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P11396"
602..1315
/codon_start=1
/product="unknown"
/protein_id="AAG17217.1"
/db_xref="GI:10441879"
/translation="MGPCITAFHSDWSSKSCAGGIGIQRVSVCLGSAALPGGG
EAGAGTGCSCPTGSRPPDRACSLGPCERTWRWTGPMGCSGSCSGTORRDIICVS
KLGTFFNTPSNCSHLPRLPALQPCQACQDRWSTPMSPCRSQCGTQREVQC
LSNTGTLSTRCPLRFSRKPNCSPQSRPDDQKSDSPHCLPLVVQARLCVPIPT
ATCCRSCHVLRSPQPS"
BASE COUNT 409 a 670 c 585 g 406 t
ORIGIN

Query Match 20.1%; Score 529; DB 9; Length 2070;
Best Local Similarity 99.8%; Pred. No. 1.7e-286;
Matches 649; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1910 ggacccaggcaacctccagcgtcaggtgcgagatccccagatgcgccgccgccccatc 1969
Db 1 GGACCCAGGCACCTCCAGCGTCAGTGCAGTCCGCCATCCGCCGCCGCCCATC 60

QY 1970 ccaggaacccctggggtctccagctcgtactggaacagagtgaggacactctgcagtct 2029
Db 61 CCAGGACACCCCTGGGGTCTCCAGTGGCTACTGGAACAGAGTGGGACACTCTGCATGCT 120

QY 2030 cagctcctcggaagagtgctggcccccatttctctgcatctcccgtagtcgg 2089
Db 121 CAGCGTCTCTGGGGAAGAGTGTCTGGCCCCCATYTTTCTCTGATCTCCCGTAGTCGG 180

QY 2090 gagaggaactggaacagcagctgtgcgagggtgccagggccccccagctccctgaac 2149

```

```

Db 181 GAGAGGAAGTGGATGAACGAGAGCTGTGCCCGGGTGCCAGGCCCGCCAGCCTCCCTGAAC 240
QY 2150 cctgcacagcaccatccccccatactggagagctggcagtgagcatctgcagcc 2209
Db 241 CCTG-CACGGCACCCCATGCCCCCATACTGGAGGCTGGCGAGTGGACATCTCTGCAGCC 299
QY 2210 gctcctgtggccccggcaccacagcaccgacagctgcagtcgagtcgaggaaggaatttggggggg 2269
Db 300 GTCCTGTGGCCCCGGCACCCAGCACCAGCTGCACTGCCGAGTGGCAGTAATTGGGGGGG 359
QY 2270 gtgctcctcgggtgccccggagcgtgtggacatctccccgggccccacatcacccag 2329
Db 360 GTGCTCTCGTGGTCCCGCGAGCGCTGTGGACATCTCCCGCGGCCACACATCACCACT 419
QY 2330 cttgcagctgcctctgtggccattgggaagtgtgctctctctggagcagtgctcg 2389
Db 420 CTTGCCAGCTGCGCCTCTGTGGCCATTGGGAAGTTGGCTCTCCTTGGAGCAGTGTCTCG 479
QY 2390 tgcggtgcggcgccggccagagagccgaggttcgctgtgttgggaacaacgtagtg 2449
Db 480 TGGGTGCGCGCGCGCGAGAGCGCGAGGTTCGCTGTGTGGGAACACACGGTGTG 539
QY 2450 aagtgcagcagcagagtgctgcagggccccccacagccccccagcagagagggcctgtg 2509
Db 540 AAGTGAGCGAGCAGGAGTGTGCTCAGGCCCGCCACAGCCCCCAGCAGAGAGGCGCTGTG 599
QY 2510 acatggggcccttactactgctgttcacagcagctggagctccaag 2559
Db 600 ACATGGGGCCCTACTACTGCTGTGTTCCACAGCGACTGGAGCTCCAAG 649

RESULT 4
AC053497
LOCUS AC053497
DEFINITION Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT
ACCESSION AC053497
VERSION AC053497.4 GI:9797901
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Waterston,R.H.
AUTHORS The sequence of Homo sapiens clone
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207815)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 12, 2000 this sequence version replaced gi:8439979.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0243G22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192323 bases at least Q40
Consensus quality: 196320 bases at least Q30
Consensus quality: 198661 bases at least Q20
Insert size: 204000; agarose-1p
Insert size: 204215; sum-of-contigs
Quality coverage: 6.36 in Q20 bases; agarose-1p

```

Quality coverage: 6.31 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1381: contig of 1381 bp in length  
\* 1382 1481: gap of unknown length  
\* 1482 2602: contig of 1121 bp in length  
\* 2603 2702: gap of unknown length  
\* 2703 4013: contig of 1311 bp in length  
\* 4014 4113: gap of unknown length  
\* 4114 5331: contig of 1218 bp in length  
\* 5332 5431: gap of unknown length  
\* 5432 6622: contig of 1191 bp in length  
\* 6623 6722: gap of unknown length  
\* 6723 8167: contig of 1444 bp in length  
\* 8168 8266: gap of unknown length  
\* 8267 9674: contig of 1408 bp in length  
\* 9675 9774: gap of unknown length  
\* 9775 10944: contig of 1169 bp in length  
\* 10944 11043: gap of unknown length  
\* 11044 12198: contig of 1155 bp in length  
\* 12199 12298: gap of unknown length  
\* 12299 14193: contig of 1895 bp in length  
\* 14194 14293: gap of unknown length  
\* 14294 15717: contig of 1424 bp in length  
\* 15718 15817: gap of unknown length  
\* 15818 16935: contig of 1118 bp in length  
\* 16936 17035: gap of unknown length  
\* 17036 18375: contig of 1340 bp in length  
\* 18376 18475: gap of unknown length  
\* 18476 19962: contig of 1487 bp in length  
\* 19963 20062: gap of unknown length  
\* 20063 21169: contig of 1107 bp in length  
\* 21170 21269: gap of unknown length  
\* 21270 22452: contig of 1183 bp in length  
\* 22453 22552: gap of unknown length  
\* 22553 24582: contig of 2030 bp in length  
\* 24583 24682: gap of unknown length  
\* 24683 26383: contig of 1701 bp in length  
\* 26384 26483: gap of unknown length  
\* 26484 27710: contig of 1227 bp in length  
\* 27711 27811: gap of unknown length  
\* 27812 29195: contig of 1385 bp in length  
\* 29196 29295: gap of unknown length  
\* 29296 30800: contig of 1505 bp in length  
\* 30801 30900: gap of unknown length  
\* 30901 32555: contig of 1655 bp in length  
\* 32556 32655: gap of unknown length  
\* 32656 34583: contig of 1998 bp in length  
\* 34584 34733: gap of unknown length  
\* 34734 37122: contig of 2369 bp in length  
\* 37123 37222: gap of unknown length  
\* 37223 39505: contig of 2284 bp in length  
\* 39506 39606: gap of unknown length  
\* 39607 41220: contig of 1614 bp in length  
\* 41221 41320: gap of unknown length  
\* 41321 42978: contig of 1658 bp in length  
\* 42979 43078: gap of unknown length  
\* 43079 45773: contig of 2695 bp in length  
\* 45774 45873: gap of unknown length  
\* 45874 55056: contig of 9183 bp in length  
\* 55057 55156: gap of unknown length  
\* 55157 67055: contig of 11903 bp in length  
\* 67056 67159: gap of unknown length  
\* 67160 80260: contig of 13101 bp in length  
\* 80261 80360: gap of unknown length  
\* 80361 96528: contig of 16168 bp in length

FEATURES

source  
1. 207815  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-243G22"  
1. 1381  
/note="assembly\_name:Contig13"  
1482. 2602  
/note="assembly\_name:Contig15"  
2703. 4013  
/note="assembly\_name:Contig21"  
4114. 5331  
/note="assembly\_name:Contig25"  
5432. 6622  
/note="assembly\_name:Contig33"  
6723. 8166  
/note="assembly\_name:Contig37"  
8267. 9674  
/note="assembly\_name:Contig38"  
9775. 10943  
/note="assembly\_name:Contig40"  
11044. 12198  
/note="assembly\_name:Contig42"  
12299. 14193  
/note="assembly\_name:Contig45"  
14294. 15717  
/note="assembly\_name:Contig47"  
15818. 16935  
/note="assembly\_name:Contig50"  
17036. 18375  
/note="assembly\_name:Contig51"  
18476. 19962  
/note="assembly\_name:Contig52"  
20063. 21169  
/note="assembly\_name:Contig53"  
21270. 22452  
/note="assembly\_name:Contig54"  
22553. 24582  
/note="assembly\_name:Contig56"  
24683. 26383  
/note="assembly\_name:Contig57"  
26484. 27710  
/note="assembly\_name:Contig58"  
27811. 29195  
/note="assembly\_name:Contig59"  
29296. 30800  
/note="assembly\_name:Contig60"  
30901. 32555  
/note="assembly\_name:Contig63"  
32656. 34653  
/note="assembly\_name:Contig64"  
34754. 37122  
/note="assembly\_name:Contig65"  
37223. 39506  
/note="assembly\_name:Contig66"  
39607. 41220  
/note="assembly\_name:Contig67"  
41321. 42978  
/note="assembly\_name:Contig68"  
43079. 45773  
/note="assembly\_name:Contig69"  
clone\_end:T7

96529 96628: gap of unknown length  
\* 96629 114529: contig of 17901 bp in length  
\* 114530 114629: gap of unknown length  
\* 114630 133673: contig of 19044 bp in length  
\* 133674 133773: gap of unknown length  
\* 133774 153791: contig of 20018 bp in length  
\* 153792 153891: gap of unknown length  
\* 153892 176079: contig of 22188 bp in length  
\* 176080 176179: gap of unknown length  
\* 176180 207815: contig of 31636 bp in length.

AC053497 4 GI:9797901  
AC053497.4 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 207815)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 207815)  
Waterston, R.H.  
Direct Submission  
Submitted (16-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Aug 12, 2000 this sequence version replaced gi:8439979.

AC053497 4 GI:9797901  
AC053497.4 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 207815)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 207815)  
Waterston, R.H.  
Direct Submission  
Submitted (16-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Aug 12, 2000 this sequence version replaced gi:8439979.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0243Q22  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-primer ET; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 192323 bases at least Q40  
Consensus quality: 196320 bases at least Q30  
Consensus quality: 198661 bases at least Q20  
Insert size: 204000; agarose-1p  
Insert size: 204215; sum-of-contigs  
Quality coverage: 6.36 in Q20 bases; agarose-1p  
Quality coverage: 6.31 in Q20 bases; sum-of-contigs

----- Project Information -----  
Center project name: H\_NH0243Q22  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-primer ET; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 192323 bases at least Q40  
Consensus quality: 196320 bases at least Q30  
Consensus quality: 198661 bases at least Q20  
Insert size: 204000; agarose-1p  
Insert size: 204215; sum-of-contigs  
Quality coverage: 6.36 in Q20 bases; agarose-1p  
Quality coverage: 6.31 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1381: contig of 1381 bp in length  
1382 1481: gap of unknown length  
1482 2602: contig of 1121 bp in length  
2603 2702: gap of unknown length  
2703 4013: contig of 1311 bp in length  
4014 4113: gap of unknown length  
4114 5331: contig of 1218 bp in length  
5332 5431: gap of unknown length  
5432 6622: contig of 1191 bp in length  
6623 8166: contig of 1444 bp in length  
8167 9674: gap of unknown length  
9675 10943: contig of 1408 bp in length  
10944 11044: gap of unknown length  
11045 12198: contig of 1155 bp in length  
12199 14193: contig of 1895 bp in length  
14194 14293: gap of unknown length  
14294 15717: contig of 1424 bp in length  
15718 15817: gap of unknown length  
15818 16935: contig of 1118 bp in length  
16936 17035: gap of unknown length  
17036 18375: contig of 1340 bp in length  
18376 18475: gap of unknown length

1 1381: contig of 1381 bp in length  
1382 1481: gap of unknown length  
1482 2602: contig of 1121 bp in length  
2603 2702: gap of unknown length  
2703 4013: contig of 1311 bp in length  
4014 4113: gap of unknown length  
4114 5331: contig of 1218 bp in length  
5332 5431: gap of unknown length  
5432 6622: contig of 1191 bp in length  
6623 8166: contig of 1444 bp in length  
8167 9674: gap of unknown length  
9675 10943: contig of 1408 bp in length  
10944 11044: gap of unknown length  
11045 12198: contig of 1155 bp in length  
12199 14193: contig of 1895 bp in length  
14194 14293: gap of unknown length  
14294 15717: contig of 1424 bp in length  
15718 15817: gap of unknown length  
15818 16935: contig of 1118 bp in length  
16936 17035: gap of unknown length  
17036 18375: contig of 1340 bp in length  
18376 18475: gap of unknown length

RESULT 5  
AC053497/c 207815 bp DNA linear HTG 13-AUG-2000  
LOCUS Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT  
DEFINITION SEQUENCE, 37 unordered pieces.





RESULT 6  
AC092479  
LOCUS  
DEFINITION  
AC092479 142023 bp DNA linear HTG 05-FEB-2002  
MUS musculus clone rp23-231115 strain C57BL/6J, WORKING DRAFT  
SEQUENCE, 3 unordered pieces.  
AC092479  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AC092479.12 GI:18497125  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
house mouse.  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 142023)  
Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.  
MUS musculus BAC Clone rp23-231115  
Unpublished  
REFERENCE  
2 (bases 1 to 142023)  
Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.  
Direct Submission  
Submitted (10-JUL-2001) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Feb 5, 2002 this sequence version replaced gi:18390246.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 32949: contig of 32949 bp in length  
\* 32950 33049: gap of unknown length  
\* 78918 79018: contig of 45869 bp in length  
\* 78919 79018: gap of unknown length  
\* 79019 142023: contig of 63005 bp in length.  
\* Location/Qualifiers  
1..142023  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="rp23-231115"  
/clone\_lib="RPCI mouse BAC library 23"  
BASE COUNT 37523 a 34071 c 31510 g 38713 t 206 others  
ORIGIN  
Query Match 1.5%; Score 39; DB 2; Length 142023;  
Best Local Similarity 100.0%; Pred. No. 7.6e-10;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2290 gagcgtgtggacatctccccggcccaacatcacccag 2328  
|||||  
Db 86350 GAGCGTGTGGACATCTCCCCGGCCCAACATCACCCAG 86388  
|||||  
RESULT 7  
AC093317  
LOCUS  
DEFINITION  
AC093317 215810 bp DNA linear HTG 29-JAN-2002  
MUS musculus clone rp23-218k6 strain C57BL/6J, WORKING DRAFT  
SEQUENCE, 10 unordered pieces.  
AC093317  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AC093317.16 GI:18390243  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
house mouse.  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 215810)  
Jia,H., Zhang,P., Lin,S. and Roe,B.A.  
MUS musculus BAC Clone rp23-218k6  
Unpublished  
2 (bases 1 to 215810)  
Jia,H., Zhang,P., Lin,S. and Roe,B.A.  
Direct Submission  
Submitted (17-AUG-2001) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Jan 29, 2002 this sequence version replaced gi:18139414.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2528: contig of 2528 bp in length  
\* 2529 2629: gap of unknown length  
\* 8258 8357: contig of 5629 bp in length  
\* 8258 8357: gap of unknown length  
\* 8358 14071: contig of 5714 bp in length  
\* 14072 14171: gap of unknown length  
\* 14172 28167: contig of 13996 bp in length  
\* 28168 28267: gap of unknown length  
\* 28268 41090: contig of 12823 bp in length  
\* 41091 41190: gap of unknown length  
\* 41191 58658: contig of 17468 bp in length  
\* 58659 58759: gap of unknown length  
\* 58759 83163: contig of 24405 bp in length  
\* 83164 83264: gap of unknown length  
\* 83264 107370: contig of 24107 bp in length  
\* 107371 107470: gap of unknown length  
\* 107471 137048: contig of 29578 bp in length  
\* 137049 137148: gap of unknown length  
\* 137149 215810: contig of 78662 bp in length.  
\* Location/Qualifiers  
1..215810  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="rp23-218k6"  
/clone\_lib="RPCI mouse BAC library 23"  
BASE COUNT 59907 a 50405 c 47367 g 57018 t 1113 others  
ORIGIN  
Query Match 1.5%; Score 39; DB 2; Length 215810;  
Best Local Similarity 100.0%; Pred. No. 7.6e-10;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2290 gagcgtgtggacatctccccggcccaacatcacccag 2328  
|||||  
Db 114431 GAGCGTGTGGACATCTCCCCGGCCCAACATCACCCAG 114469  
|||||  
RESULT 8  
BC016215  
LOCUS  
DEFINITION  
BC016215 1409 bp mRNA linear ROD 05-NOV-2001  
MUS musculus, Similar to hypothetical protein FLJ13710, clone  
MGC:28749 IMAGE:4482484, mRNA, complete cds.  
AC093317  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BC016215.1 GI:16740678  
MGC.  
house mouse.  
MUS musculus



mus musculus clone RP23-176010, LOW-PASS SEQUENCE SAMPLING.

AC101198  
AC101198\_1 GI:17059962  
HTG; HGSC\_PHASE0.  
house mouse.

Mus musculus  
Mus musculus  
Mus musculus

(bases 1 to 62385)  
Linton,B., Linton,L., Nusbaum,C. and Lander,E.  
unpublished  
2 (bases 1 to 62385)  
Linton,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukigalter,B.,  
Campione,A., Camarata,J., Campionaro,A., Chang,J., Chazaro,B.,  
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galaan,J., Gardyna,S.,  
Grande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hafago,S., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,K., Liu,G.,  
MacDonald,P., Major,J., Marquis,N., Matthews,C.,  
McEwan,P., McKernan,K., McEwan,J., McPheters,R., Melrim,J.,  
Mihova,T., Mienna,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnuppach,R.,  
Seaman,S., Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Thompson,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information -----  
Center project name: L16125  
Center clone name: 176\_Q10  
-----

\* NOTE: This record contains 80 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

I 692: contig of 692 bp in length  
693 792: gap of 100 bp  
793 1503: contig of 711 bp in length  
1504 1603: gap of 100 bp  
1604 2312: contig of 709 bp in length  
2313 2412: gap of 100 bp  
2413 3068: contig of 656 bp in length  
3069 3168: gap of 100 bp  
3169 3846: contig of 678 bp in length  
3847 3946: gap of 100 bp  
3947 4630: contig of 684 bp in length  
4631 4730: gap of 100 bp

\* 1 602: contig of 692 bp in length  
 \* 792: gap of 100 bp  
 \* 1503: contig of 711 bp in length  
 \* 1603: gap of 100 bp  
 \* 2312: contig of 709 bp in length  
 \* 2412: gap of 100 bp  
 \* 3068: contig of 656 bp in length  
 \* 3168: gap of 100 bp  
 \* 3846: contig of 678 bp in length  
 \* 3946: gap of 100 bp  
 \* 4630: contig of 684 bp in length  
 \* 4730: gap of 100 bp

\* 4731 5409: contig of 679 bp in length  
\* 5410 5509: gap of 100 bp  
\* 5510 6217: contig of 708 bp in length  
\* 6218 6317: gap of 100 bp  
\* 6318 6972: contig of 655 bp in length  
\* 6973 7072: gap of 100 bp  
\* 7073 7767: contig of 695 bp in length  
\* 7768 7867: gap of 100 bp  
\* 7868 8572: contig of 705 bp in length  
\* 8573 8672: gap of 100 bp  
\* 8673 9377: contig of 705 bp in length  
\* 9378 9477: gap of 100 bp  
\* 9478 10144: contig of 667 bp in length  
\* 10145 10244: gap of 100 bp  
\* 10245 10883: contig of 639 bp in length  
\* 10884 10983: gap of 100 bp  
\* 10984 11665: contig of 682 bp in length  
\* 11666 11765: gap of 100 bp  
\* 11766 12433: contig of 668 bp in length  
\* 12434 12533: gap of 100 bp  
\* 12534 13214: contig of 681 bp in length  
\* 13215 13314: gap of 100 bp  
\* 13315 14007: contig of 693 bp in length  
\* 14008 14107: gap of 100 bp  
\* 14108 14811: contig of 704 bp in length  
\* 14812 14911: gap of 100 bp  
\* 14912 15614: contig of 703 bp in length  
\* 15615 15714: gap of 100 bp  
\* 15715 16428: contig of 714 bp in length  
\* 16429 16528: gap of 100 bp  
\* 16529 17220: contig of 692 bp in length  
\* 17221 17320: gap of 100 bp  
\* 17321 17979: contig of 659 bp in length  
\* 17980 18079: gap of 100 bp  
\* 18080 18759: contig of 680 bp in length  
\* 18760 18859: gap of 100 bp  
\* 18860 19526: contig of 667 bp in length  
\* 19527 19626: gap of 100 bp  
\* 19627 20303: contig of 677 bp in length  
\* 20304 20403: gap of 100 bp  
\* 20404 21090: contig of 687 bp in length  
\* 21091 21190: gap of 100 bp  
\* 21191 21897: contig of 707 bp in length  
\* 21898 21997: gap of 100 bp  
\* 21998 22668: contig of 671 bp in length  
\* 22669 22768: gap of 100 bp  
\* 22769 23467: contig of 699 bp in length  
\* 23468 23567: gap of 100 bp  
\* 23568 24257: contig of 690 bp in length  
\* 24258 24357: gap of 100 bp  
\* 24358 25045: contig of 688 bp in length  
\* 25046 25145: gap of 100 bp  
\* 25146 25810: contig of 665 bp in length  
\* 25811 25910: gap of 100 bp  
\* 25911 26573: contig of 663 bp in length  
\* 26574 26673: gap of 100 bp  
\* 26674 27341: contig of 668 bp in length  
\* 27342 27441: gap of 100 bp  
\* 27442 28129: contig of 688 bp in length  
\* 28130 28229: gap of 100 bp  
\* 28230 28907: contig of 678 bp in length  
\* 28908 29007: gap of 100 bp  
\* 29008 29697: contig of 690 bp in length  
\* 29698 29797: gap of 100 bp  
\* 29798 30503: contig of 706 bp in length  
\* 30504 30603: gap of 100 bp  
\* 30604 31303: contig of 700 bp in length  
\* 31304 31403: gap of 100 bp  
\* 31404 32047: contig of 644 bp in length  
\* 32048 32147: gap of 100 bp  
\* 32148 32833: contig of 686 bp in length  
\* 32834 32933: gap of 100 bp  
\* 32934 33615: contig of 682 bp in length

\* 33616 33715: gap of 100 bp  
\* 33716 34416: contig of 701 bp in length  
\* 34417 34516: gap of 100 bp  
\* 34517 35188: contig of 673 bp in length  
\* 35190 35289: gap of 100 bp  
\* 35290 35952: contig of 663 bp in length  
\* 35953 36052: gap of 100 bp  
\* 36053 36710: contig of 658 bp in length  
\* 36711 36810: gap of 100 bp  
\* 36811 37486: contig of 676 bp in length  
\* 37487 37586: gap of 100 bp  
\* 37587 38276: contig of 690 bp in length  
\* 38277 38376: gap of 100 bp  
\* 38377 39089: contig of 713 bp in length  
\* 39090 39189: gap of 100 bp  
\* 39190 39825: contig of 636 bp in length  
\* 39826 39925: gap of 100 bp  
\* 39926 40531: contig of 606 bp in length  
\* 40532 40631: gap of 100 bp  
\* 40632 41161: contig of 530 bp in length  
\* 41162 41261: gap of 100 bp  
\* 41262 41931: contig of 670 bp in length  
\* 41932 42031: gap of 100 bp  
\* 42032 42699: contig of 668 bp in length  
\* 42700 42799: gap of 100 bp  
\* 42800 43474: contig of 675 bp in length  
\* 43475 43574: gap of 100 bp  
\* 43575 44272: contig of 698 bp in length  
\* 44273 44372: gap of 100 bp  
\* 44373 45069: contig of 697 bp in length  
\* 45070 45169: gap of 100 bp  
\* 45170 45838: contig of 667 bp in length  
\* 45837 45938: gap of 100 bp  
\* 45937 46642: contig of 705 bp in length  
\* 46642 46741: gap of 100 bp  
\* 46742 47447: contig of 706 bp in length  
\* 47448 47547: gap of 100 bp  
\* 47548 48248: contig of 702 bp in length  
\* 48250 48349: gap of 100 bp  
\* 48350 49000: contig of 651 bp in length  
\* 49001 49100: gap of 100 bp  
\* 49101 49777: contig of 677 bp in length  
\* 49778 49877: gap of 100 bp  
\* 49878 50555: contig of 678 bp in length  
\* 50556 50655: gap of 100 bp  
\* 50656 51344: contig of 689 bp in length  
\* 51345 51444: gap of 100 bp  
\* 51445 52140: contig of 686 bp in length  
\* 52141 52240: gap of 100 bp  
\* 52241 52926: contig of 686 bp in length  
\* 52927 53026: gap of 100 bp  
\* 53027 53732: contig of 706 bp in length  
\* 53733 53832: gap of 100 bp

Query Match 0.9%; Score 23; DB 2; Length 62385;  
Best Local Similarity 100.0%; Pred.No. 0.81;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 170 ccagccctcgagggtgggggtg 192  
|||||  
Db 51295 CCAGCCCTCGAGGCTGGGGGTG 51273

RESULT 12  
AC026636/c  
LOCUS AC026636 151333 bp DNA linear HTG 27-AUG-2001  
DEFINITION Homo sapiens chromosome 15 clone RP11-170E16 map 15, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 4 unordered pieces.  
ACCESSION AC026636  
VERSION AC026636.5 GI:5291088  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 151333)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 15, clone RP11-170E16

Unpublished

2 (bases 1 to 151333)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castelle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Horton,L., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,K., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pischner,B., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 26, 2001 this sequence version replaced gi:15148317.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8704

Center clone name: 170\_E\_16

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 25339: contig of 25339 bp in length

2 25340 25439: gap of 100 bp

3 25440 36562: contig of 11123 bp in length

4 36563 36662: gap of 100 bp

5 36663 61806: contig of 25144 bp in length

6 61807 61906: gap of 100 bp

7 61907 151333: contig of 89427 bp in length.

----- Location/Qualifiers

1. 151333

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="15"

/map="15"

/clone="RP11-170E16"

/clone\_lib="RPC1-11 Human Male BAC"

BASE COUNT 44563 a 32212 c 31615 g 42618 t 325 others

ORIGIN

Query Match 0.98; Score 23; DB 2; Length 151333;

Best Local Similarity 100.0%; Pred. No. 0.82;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1600 tccatcatcaatgggaactgggc 1622

|||||

DB 32392 tccatcatcaatgggaactgggc 32370

RESULT 13

AC104938/c

LOCUS

DEFINITION

Homo sapiens chromosome 15 clone CTB-2524L6 map 15, WORKING DRAFT

SEQUENCE, 7 ordered pieces.

AC104938

AC104938.1 GI:17977316

HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 184483)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 15, clone CTB-2524L6

Unpublished

2 (bases 1 to 184483)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L21457

Center clone name: 2524\_L\_6

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 182343 bases at least Q40

Consensus quality: 182981 bases at least Q30

Consensus quality: 183229 bases at least Q20

Insert size: 183883; sum-of-contigs

Quality coverage: 10.5 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 549: contig of 549 bp in length

\* 550 649: gap of 100 bp  
\* 14178 14177: contig of 13528 bp in length  
\* 14178 14277: gap of 100 bp  
\* 22038 22037: contig of 7760 bp in length  
\* 22038 22137: gap of 100 bp  
\* 22138 40150: contig of 18013 bp in length  
\* 40151 40250: gap of 100 bp  
\* 40251 66123: contig of 25873 bp in length  
\* 66124 66223: gap of 100 bp  
\* 66224 114326: contig of 48103 bp in length  
\* 114327 114426: gap of 100 bp  
\* 114427 184483: contig of 70057 bp in length.

FEATURES

source  
1. 184483  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone\_lib="CIRB Human BAC"  
1. 549  
/note="assembly\_fragment"  
550. 14177  
/note="assembly\_fragment"  
14278. 22037  
/note="assembly\_fragment"  
22138. 40150  
/note="assembly\_fragment"  
40251. 66123  
/note="assembly\_fragment"  
66224. 114326  
/note="assembly\_fragment"  
114427. 184483  
/note="assembly\_fragment"  
BASE COUNT 55127 a 39050 c 39152 g 50554 t 600 others  
ORIGIN

Query Match 0.9%; Score 23; DB 2; Length 184483;  
Best Local Similarity 100.0%; Pred. NO. 0.82;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1600 tccatcatcatgggaactgggc 1622  
Db 154963 TCCATCATCAATGGGAACCTGGC 154941

RESULT 14

AC019036/c  
LOCUS 168872 bp DNA linear HTG 08-APR-2000  
DEFINITION Homo sapiens chromosome 4 clone RP11-12318 map 4, WORKING DRAFT  
SEQUENCE, 20 unordered pieces.

AC019036  
AC019036 GI:7528071  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS human.  
SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 168872)  
Birken, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 4, clone RP11-12318  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 168872)  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,

Boquslavkiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Derebellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,  
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (30-DEC-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 8, 2000 this sequence version replaced g1:6730845.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WBIR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4941  
Center clone name: 123.1.8  
----- Summary Statistics  
Sequencing Vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 156691 bases at least Q40  
Consensus quality: 162047 bases at least Q30  
Consensus quality: 164542 bases at least Q20  
Insert size: 166000; agarose-fp  
Insert size: 166972; sum-of-contigs  
Quality coverage: 4.1 in Q20 bases; agarose-fp  
Quality coverage: 4.1 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1092: contig of 1092 bp in length  
\* 1093 1192: gap of 100 bp  
\* 1193 4028: contig of 2836 bp in length  
\* 4029 4128: gap of 100 bp  
\* 4129 8354: contig of 4226 bp in length  
\* 8355 8454: gap of 100 bp  
\* 8455 11747: contig of 3293 bp in length  
\* 11748 11847: gap of 100 bp  
\* 11848 16091: contig of 4244 bp in length  
\* 16092 16191: gap of 100 bp  
\* 16192 19148: contig of 2957 bp in length  
\* 19149 19248: gap of 100 bp  
\* 19249 25597: contig of 6349 bp in length  
\* 25598 25697: gap of 100 bp  
\* 25698 31571: contig of 5874 bp in length  
\* 31572 31671: gap of 100 bp  
\* 31672 38194: contig of 6523 bp in length  
\* 38195 38294: gap of 100 bp  
\* 38295 44033: contig of 5739 bp in length  
\* 44034 44133: gap of 100 bp  
\* 44134 51560: contig of 7427 bp in length  
\* 51561 51660: gap of 100 bp  
\* 51661 60691: contig of 9031 bp in length

\* 60592 60791: gap of 100 bp  
\* 60792 69942: contig of 9151 bp in length  
\* 69943 70042: gap of 100 bp  
\* 70043 77752: contig of 7710 bp in length  
\* 77753 77852: gap of 100 bp  
\* 77853 89028: contig of 11176 bp in length  
\* 89029 89128: gap of 100 bp  
\* 89129 99478: contig of 10350 bp in length  
\* 99479 99578: gap of 100 bp  
\* 99579 115864: contig of 16286 bp in length  
\* 115865 115964: gap of 100 bp  
\* 115965 134124: contig of 18160 bp in length  
\* 134125 134224: gap of 100 bp  
\* 134225 152621: contig of 18397 bp in length  
\* 152622 152721: gap of 100 bp  
\* 152722 168872: contig of 16151 bp in length.

FEATURES

source  
1. .168872  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-12318"  
/clone\_lib="RP11-11 Human Male BAC"  
misc\_feature  
1. .1092  
/note="assembly\_fragment"  
1193. .4028  
/note="assembly\_fragment"  
4129. .8354  
/note="assembly\_fragment"  
8455. .11747  
/note="assembly\_fragment"  
11848. .16091  
/note="assembly\_fragment"  
16192. .19148  
/note="assembly\_fragment"  
19249. .25597  
/note="assembly\_fragment"  
25698. .31571  
/note="assembly\_fragment"  
31672. .38194  
/note="assembly\_fragment"  
38295. .44033  
/note="assembly\_fragment"  
44134. .51560  
/note="assembly\_fragment"  
51661. .60691  
/note="assembly\_fragment"  
60792. .69942  
/note="assembly\_fragment"  
70043. .77752  
/note="assembly\_fragment"  
77853. .89028  
/note="assembly\_fragment"  
89129. .99478  
/note="assembly\_fragment"  
99579. .115864  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
115965. .134124  
/note="assembly\_fragment"  
134225. .152621  
/note="assembly\_fragment"  
152722. .168872  
/note="assembly\_fragment"  
BASE COUNT 52704 a 30458 c 31011 g 52799 t 1900 others  
ORIGIN

Query Match 0.8%; Score 22; DB 2; Length 168872;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 tgcccttgcattgccactgca 506  
|||||  
Db 115680 TGCCCTTGCATTGCCACTGCA 115659

RESULT 15

LOCUS AC097488 175970 bp DNA linear PRI 09-FEB-2002  
DEFINITION Homo sapiens chromosome 4 clone RP11-218C23, complete sequence.  
ACCESSION AC097488 AC067828  
VERSION AC097488.2 GI:18642911  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 175970)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (18-OCT-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 3 (bases 1 to 175970)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT

On Feb 9, 2002 this sequence version replaced gi:16259095.

FEATURES

source  
1. .175970  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="RP11-218C23"  
BASE COUNT 50188 a 32373 c 33188 g 60221 t  
ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 175970;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 tgcccttgcattgccactgca 506  
|||||

Db 133569 TGCCCTTGCATTGCCACTGCA 133590

Search completed: July 23, 2002, 20:20:02  
Job time: 9842 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 13:05:10 ; Search time 3244.83 Seconds  
(without alignments)  
16987.176 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 attgagaactgactgcag.....cctcagccttcagcatag 2634

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb\_ba.\*
- 2: gb\_hcg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	2557.4	97.1	4230	6	AX342636
2	878.2	33.3	176550	9	AL356356
3	854.2	32.4	207815	2	AC053497
4	636	24.2	2070	9	AF217974
5	251.4	9.5	142023	2	AC092479
6	250.2	9.5	207815	2	AC053497
7	239.4	9.1	215810	2	AC093317
8	211.6	8.0	2703	9	BC008840
9	165.2	6.3	2964	9	AK023772
10	163.4	6.2	4271	10	AF314171
11	113.2	4.3	2805	6	AX319854
12	106.8	4.1	3675	6	AX319852
13	99	3.8	2410	6	AX342638
14	97	3.7	86446	2	AC014412
15	97	3.7	182525	3	AC009213
16	97	3.7	239171	3	AE003765
17	96.6	3.7	5061	6	E55365
18	96.2	3.7	5353	6	AX128433
19	94	3.6	2848	9	AF140674
20	93	3.5	11008	3	AF078161
21	92.2	3.5	7100	3	AF205357
22	91.6	3.5	3312	6	AX319851
23	91.6	3.5	3312	6	E55273
24	91.4	3.5	3132	6	AX327754
25	91.4	3.5	3132	6	AX111571
26	91.4	3.5	3329	6	AX327752
27	88.4	3.4	3400	9	AF163762
28	84.8	3.2	125020	9	AF429315
29	80.6	3.1	4287	6	AX319855
30	78	3.0	125020	9	AF429315
31	77.4	2.9	1119	6	AX224813
32	77.4	2.9	1953	6	AX224817
33	77.4	2.9	2316	6	AX224821
34	77.4	2.9	4854	6	AX224823
35	77.2	2.9	3179	9	AF140675
36	74.2	2.8	4067	9	AB011177
37	74.2	2.8	5199	3	AF043121
38	73.4	2.8	2445	6	AX327757
39	73.4	2.8	3207	6	AX327747
40	73.4	2.8	3402	6	AX338539
41	73.4	2.8	3403	6	AX327745
42	73.4	2.8	3403	6	AX338537
43	73.4	2.8	3471	6	AX327755
44	73	2.8	1341	6	AX224811
45	73	2.8	2175	6	AX224815

ALIGNMENTS

RESULT	1	AX342636	Sequence 33 from Patent WO0198468.	4230 bp	DNA	linear	PAT 12-JAN-2002
LOCUS	AX342636	Sequence 33 from Patent WO0198468.					
DEFINITION	AX342636						
ACCESSION	AX342636.1	GI:18152033					
VERSION							
KEYWORDS							
SOURCE		human.					
ORGANISM		Homo sapiens					
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS		Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J., Tribouley, C.M., Deleage, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A., Habouley, A., Khan, F.A., Walia, N.K., Yao, M.G., Lu, D.A., Patterson, C., Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R. Patent: WO 0198468-A 33 27-DEC-2001; Incyte Genomics, Inc. (US)					
JOURNAL							
FEATURES		Location/Qualifiers					
source		1. 4230					
		/organism="Homo sapiens"					
		/db xref="taxon.9606"					
		/note="Incyte ID No: 7604035CB1"					

[illegible]

Db 2268 GGAAGAGGTGTGGCGGCCCATTTTCTCTGCATCTCCCGTAGTGGGAGAGGAACGTG 2327  
 QY 2101 gatgaacacacgtatgcgcgggtgacagcccccagcctccctgaacctgcacagc 2160  
 Db 2328 GATGAACACAGTGTGGCGGGTGCCAGGCCCCAGCCTCCCTTGACCCCTGCACAGGC 2387  
 QY 2161 accccatgccccccatactggaggtgcggagtggaacatcctgcagccgtcctgtggc 2220  
 Db 2388 ACCCCATGCCCCCATACTGGAGGCTGGCGAGTGGACATCTCAGCGCGCTCCTGTGGC 2447  
 QY 2221 cccggcaccagcaccgacacgtcagtgccggcaggaattggggggtggtctctcg 2280  
 Db 2448 CCGGACACCCAGCACCGCCAGTGCAGTGCAGGCGCGGAGGAATTTGGGGGGGTGCTCTCG 2507  
 QY 2281 gtgccccggagcgtgtggacatctcccccggcccaacatcacccagcttggcagctg 2340  
 Db 2508 GTGCCCCGGAGCGCTGTGGACATCTCCCGGCCCAACATCACCGAGCTTGCACAGTG 2567  
 QY 2341 cgcctctgtgcccattgggaagtgtgctctccttggagccagtgctcogtgcggtggc 2400  
 Db 2568 CGCCTCTGTGCCCATTTGGGAAGTTGGCTCTCTTGGAGCCAGTGTCCGTGCGTGGCG 2627  
 QY 2401 cggggccagagaagccgcaggttcgtgtgtgggaacacggtgatgaagtgcagcag 2460  
 Db 2628 CCGGCCACAGAGACCGCCAGCTTCGTGTGGGAACAACGGTGATGAAGTGAAGCAG 2687  
 QY 2461 gaggagtggtgcagggccccccagccccccagcagagagggcgtgacatggggccc 2520  
 Db 2688 CAGGAGTGTGGTCAAGGCCGCCACAGCCCCCAGCAGAGAGGGCTGTGACATGGGGCCC 2747  
 QY 2521 ttactactgctgttccacagcactggagctccaa 2559  
 Db 2748 TGTACTACTGCTGTTCACAGCCAGCTGGAGCTCCAAG 2786

RESULT 2  
 LOCUS AL356356 176550 bp DNA linear PRI 30-JAN-2002  
 DEFINITION Human DNA sequence from clone Rp11-54A4 on chromosome 1, complete sequence.  
 ACCESSION AL356356  
 VERSION AL356356.17 GI:18476570  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 176550)  
 Direct Submission  
 Heath,P.  
 Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humqueresanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Feb 1, 2002 this sequence version replaced gi:18250748.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>  
 Rp11-54A4 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACE3.6  
 IMPORTANT: This sequence is not the entire insert of clone Rp11-54A4 it may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true left end of clone Rp11-54A4 is at 1 in this sequence. The true left end of clone Rp11-363122 is at 174551 in this sequence.

FEATURES  
 Location/Qualifiers  
 1..176550  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="Rp11-54A4"  
 /clone\_lib="RPCI-11.1"  
 misc\_feature  
 50307..50323  
 /note="Single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."  
 123022..123038  
 /note="Sequence from uni-directional primer reads and dGTP big dye terminator reads only."  
 BASE COUNT 45999 a 42218 c 42262 g 46071 t  
 ORIGIN

Query Match 33.3%; Score 878.2; DB 9; Length 176550;  
 Best Local Similarity 84.3%; Pred. No. 2.6e-173;  
 Matches 1077; Conservative 0; Mismatches 28; Indels 172; Gaps 1;

QY 53 tccctcagctctgttggtatcaggaggtgtgttcggagacactctcttcagacacctacag 112  
 Db 80134 TCTCTCTGACCTTTTCTCTATATAGGTGTGTTCGGACACTCTCTTCAGACACTACAG 80193  
 QY 113 agggagcccgcccgaggtgtgtggtgacactggttcagtggtggtctctgtctccc 172  
 Db 80194 AGGAGGCCAGGGCCCGAAGGTGTCTGGGACCTTGGTCCAGTGGCCCTCTGTCTCC 80253  
 QY 173 agccctgcgggtgggggtgcagcgcaggcggagcgtcagctcctcagtcagtcagcgc 232  
 Db 80254 AGCCTCGGGGTGGGGTGCAGCGCAGAGAGCCGACATGTCAGCTCCCTACAGTGCGAG 80313  
 QY 233 tccaccagctctccctccctccctcccgcccccagacacacacagcctcctccccc 292  
 Db 80314 TCCACCGAGTCTGCCCTCTCCCTCCCGCCCCCAAGACATCCAGAGCCCTCTCTCCCC 80373  
 QY 293 ggggccagggtcccgagaccccgagacttctccagaaacctccctctgtacagacacagt 352  
 Db 80374 GGGCCAGGGTCCAGACCCAGAGACTTCTCCAGAAACCTCCCTTGTACAGGACACAGT 80433  
 QY 353 ctcggggaagggtggccaccttcagagtcgcgttccacacctaggagagagagagacc 412  
 Db 80434 CTCGGGAAGGGGTGGCCACTTCGAGTCCGCTTCCACCTCCAGGAGAGAGAGACC 80493  
 QY 413 aggagattcagcggccagg ----- 432  
 Db 80494 AGGAGATTCGAGCGGCCAGGAGGTGAGAGGCTGGGTGAAGAGGTGGCTTGGGCAAG 80553  
 QY 433 ----- 432  
 Db 80554 GTGGGGCTTAGCATGAAGGAAGGAGGACGACGGTGGCTTCAGAGGGCTTGTCTTTGGT 80613  
 QY 433 ----- 432  
 Db 80614 GCCAGGAAGTAACCTGGGAACAGGGTGCAGAGTGTGTCAGAGCTGTGCCCCACATCTG 80673  
 QY 433 ----- aggtcccggtctcgagaccccatcaagccaggaatgttcggttatggg 480  
 Db 80674 AAACACCTTCTCAGGTCCCGCTTCGAGACCCCATCAAGCCAGGAATGTTTCGGTTATGGG 80733











```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 207815)
Waterston,R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 207815)
Waterston,R.H.
Direct Submission
Submitted (16-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 12, 2000 this sequence version replaced gi:8439979.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: R_NH0243G22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192323 bases at least Q40
Consensus quality: 196320 bases at least Q30
Consensus quality: 198661 bases at least Q20
Insert size: 204000; agarose-fp
Insert coverage: 6.36 in Q20 bases; agarose-fp
Quality coverage: 6.31 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1381: contig of 1381 bp in length
* 1382
* 1481: gap of unknown length
* 1482
* 2602: contig of 1121 bp in length
* 2603
* 2702: gap of unknown length
* 2703
* 4013: contig of 1311 bp in length
* 4014
* 4113: gap of unknown length
* 4114
* 5331: contig of 1218 bp in length
* 5332
* 5431: gap of unknown length
* 5432
* 6622: contig of 1191 bp in length
* 6623
* 8166: contig of 1444 bp in length
* 8167
* 8266: gap of unknown length
* 8267
* 9574: contig of 1408 bp in length
* 9575
* 10943: contig of 1169 bp in length
* 10944
* 11043: gap of unknown length
* 11044
* 12198: contig of 1155 bp in length
* 12199
* 12298: gap of unknown length
* 12299
* 14193: contig of 1895 bp in length
* 14194
* 14293: gap of unknown length
* 14294
* 15717: contig of 1424 bp in length
* 15718
* 15817: gap of unknown length
* 15818
* 16935: contig of 1118 bp in length
* 16936
* 17035: gap of unknown length
* 17036
* 18375: contig of 1340 bp in length
* 18376
* 18475: gap of unknown length
* 18476
* 19562: contig of 1487 bp in length
* 19563
* 20062: gap of unknown length
*
* 20063
* 21170
* 21269: gap of unknown length
* 21270
* 22452: contig of 1183 bp in length
* 22453
* 22552: gap of unknown length
* 22553
* 24582: contig of 2030 bp in length
* 24583
* 24682: gap of unknown length
* 24683
* 26383: contig of 1701 bp in length
* 26384
* 26483: gap of unknown length
* 26484
* 27711
* 27810: contig of 1227 bp in length
* 27811
* 29195: contig of 1385 bp in length
* 29196
* 29295: gap of unknown length
* 29296
* 30800: contig of 1505 bp in length
* 30801
* 30900: gap of unknown length
* 30901
* 32555: contig of 1655 bp in length
* 32556
* 32655: gap of unknown length
* 32656
* 34653: contig of 1998 bp in length
* 34654
* 37122: contig of 2369 bp in length
* 37123
* 37222: gap of unknown length
* 37223
* 39506: contig of 2284 bp in length
* 39507
* 39606: gap of unknown length
* 39607
* 41220: contig of 1614 bp in length
* 41221
* 41320: gap of unknown length
* 41321
* 42978: contig of 1658 bp in length
* 42979
* 43078: gap of unknown length
* 43079
* 45773: contig of 2695 bp in length
* 45774
* 45873: gap of unknown length
* 45874
* 55136: gap of unknown length
* 55137
* 67059: contig of 11903 bp in length
* 67060
* 67159: gap of unknown length
* 67160
* 80260: contig of 13101 bp in length
* 80261
* 80360: gap of unknown length
* 80361
* 96528: contig of 16168 bp in length
* 96529
* 96628: gap of unknown length
* 96629
* 114539: contig of 17901 bp in length
* 114540
* 114629: gap of unknown length
* 114630
* 133673: contig of 19044 bp in length
* 133674
* 133773: gap of unknown length
* 133774
* 153791: contig of 20018 bp in length
* 153792
* 153891: gap of unknown length
* 153892
* 176080: contig of 22188 bp in length
* 176081
* 176179: gap of unknown length
* 176180
* 207815: contig of 31636 bp in length.
*
* FEATURES
* source
* 1. .207815
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="1"
* /clone="Rp11-243G22"
*
* misc_feature
* 1. .1381
* /note="assembly_name:Contig13"
* 1482. .2602
* /note="assembly_name:Contig15"
* 2703. .4013
* /note="assembly_name:Contig21"
* 4114. .5331
* /note="assembly_name:Contig25"
* 5432. .6622
* /note="assembly_name:Contig33"
* 6723. .8166
* /note="assembly_name:Contig37"
* 8267. .9574
* /note="assembly_name:Contig38"
* 9575. .10943
* /note="assembly_name:Contig40"
* 11044. .12198
* /note="assembly_name:Contig42"
* 12299. .14193
* /note="assembly_name:Contig45"
* 14294. .15717
* /note="assembly_name:Contig47"

```





[illegible]

YVIMGTNAISQVPHRRPFPFNGVMTGRSQEAGEQKGRNEKEDLRGEAFEMFT  
SESATQTPVRHPRDPRNDNLVPPAPPPRRSRDHNLKOLGTTECTSTCGKGSQYP  
IFRCVHSTHEAESVSDSMKPTPEECNLPFCFWDIGEMSCSTCGKGLGMOH  
QVLCIKQVANSITVDPYRCOHLKLETPETSTCOLKICSEWQIFDTWTSVPCGVGO  
RTDVKVNSITVDVDEECNMKLPNDIENCCPKAKSWFLPENSERSAECGAGY  
RTSVVCMNHVSFLPLEGGNNRPAETPCDNGPCTGKVEWFAQSWQSCIEGSGOT  
QOREVTCVKNADTFVLDPSECFLEKPPSQOCHLPCGAKFSTEWMSCKSGOG  
GFYREVRCLSDMTLSNLDPPQLKPERESCNDQCVPEYDENCKDKYNCNVVVOA  
RLCVNYKTKACCASCVRANRGTGFLGR"

BASE COUNT 705 a 862 c 796 g 601 t  
ORIGIN

Query Match 6.3%; Score 165.2; DB 9; Length 2964;

Best Local Similarity 56.1%; Pred. No. 2.4e-24;

Mismatches 258; Indels 3; Gaps 1;

Matches 333; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 1960 cggccccctccaggagacacccctgggtctccagctggtactgtaaaacagagtgagacac 2019

DB 474 CCACAGACCCGAGCCGCCAGCGGCGGATCACAATGGAAGCAGCTTGGGACA 533

QY 2020 tctgcatgctcagctcctgcgggaagtgctggtgccccatttctctgcatctcc 2079

DB 534 ACAGATGTTCCAGGACCTGTGGAAAGATCGCAGTACCTATTTCGCTGTGTGCAC 593

QY 2080 cgtgagtcggagaggaactgagatgaacagcagctgtgcgggtgcccagccacac 2139

DB 594 AGAAGCAGTATGAAGAGGCTCTGAGATTTACTGTGATCCAGCATGAAGCCGACCC 653

QY 2140 tccctgaacctgcacggcagccctacccctacccctacccctacccctacccctaccc 2199

DB 654 GAGGAGGACCTGCAACATCTTCCTTGCACGCTTCTGGGACATCGGGAGTGTCT 713

QY 2200 tctgagcgcctctgtggtgccccagcagcagcagcagcagcagcagcagcagcagc 2259

DB 714 GAGTGAGCAAGACCTGTGCGCTGGCGATGACACCGCAGGCTTGTGCGCGCAGG 773

QY 2260 ttgggggggggtgctcctcctggtgccccggagcgtgtggacatctccccggcccaac 2319

DB 774 TAGCCCAACCGACCTGAGCGGTGACGCTTACCCTGCCAGCACCTGGAGAACCTGAG 833

QY 2320 ataccagctcttgccagctgctctgtgccccctgtgccccctgtgccccctgtgcccc 2379

DB 834 ACCACGACACCTGCGCACTCAAGATGTCAGGAGTGCAGATCCGACCGACCTGAC 893

QY 2380 cagtcctcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2439

DB 894 TCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 953

QY 2440 aacggtgatgaagtgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2499

DB 954 ATTGGGGATGTGGTACGATGAGGAATGACATGAAGCTC---CGGCGCAATGACATT 1010

QY 2500 gaggcctgacatggcggcctactactgctgctgctgctgctgctgctgctgctgctg 2553

DB 1011 GAGAACTGCGACATGGAGACCTGTGCGAGAGAGTGTGCTTCTACCGAGTGGAGC 1064

RESULT 10

AF314171

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 4271)

Kramerova, I.A., Kawaguchi, N., Nelson, R.E., Fessler, J.H., Chen, Y.,

Sieron, A.L., Prockop, D.J., and Fessler, J.H.

AF314171 Mus musculus papilin mRNA, complete cds. ROD 21-DEC-2000

DEFINITION Mus musculus papilin mRNA, complete cds.

ACCESSION AF314171

VERSION AF314171.1 GI:11935121

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 4271)

Kramerova, I.A., Kawaguchi, N., Nelson, R.E., Fessler, J.H., Chen, Y.,

Sieron, A.L., Prockop, D.J., and Fessler, J.H.

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

1. .4271

/organism="Mus musculus"

/strain="CD-1"

/db\_xref="taxon:10090"

/dev\_stage="embryo"

63. .3905

/note="secreted; similar to Drosophila melanogaster

extracellular matrix protein papilin"

/codon\_start=1

/product="papilin"

/protein\_id="AAG41980.1"

/db\_xref="GI:11935122"

/translation="MQLFLLFLLSTPGSWARNVRQSDTWGTWGEWSPCSRCTG

GGISFRPCYSQRDGGTSCVGPAPSHRTCHTESCPDGVRFQADQCAFEFGDTDFQ

RRYRWLPYAAPNKCLNCPKGNFYKHDAVDGTCEPGORDICVDGVGVVGC

DKLSDTKQEDKCLQCGDSSCYPTGTFDGNDLSRGYNQIFLIPAGATGIRIEPA

ASRNFLAVSIRGEYILNGHWTTEAQAALPVASTVLOYRGEVGEDLAPERLQAGT

EPVILLISQESNPGVHYEYILPANDPGRFSPWSHGSDCSAECGGGQGRVLFCTI

DNAYPDHMCQHPRPTRRSCNTQPKTKRWKVPWTPCVSCGGGVQSVRSYCLIS

SDGTGQEAEEATCAGLAKPTTQACNLQCAVMSVEFWGECSTVCTGIRKRSVT

CRDGEQSPVHAACLLKQDPTLTPCQVACPVFQGAHVMSWLSKSCSGSIRRR

QVVTIGPPGRCVDLQSSKPAEMACNRQPCPLQEVPSIQDPTRSSDPRILSGPRV

SPVSDGECQWAPLERPRAONREGQDNLSSAGRAPTLQRPPOHPPLRPSGPRDC

RHPHGCGPGHTPSLPGQWQCPLAGACLSQRYGCCPDGVSAAAGPOAGTSHG

SNTNRPGRVASKNPKIHQFOAHEFESEKSNFGCCYCNVSAAGPAGLEGCVG

QPSYAPVRCLLPSAQCGDMAWRTYFASVGRNFRFVGGCHGANNFASGECMN

TCGQHGPRRPAAGRAHRYDGGQGGQGGQEPDWHAGATIPRLPSGSGPWRRE

QEPAGPEPHIPAYGNRPGQETRPVPLDREARPAVPTSPSRILRAGSEPLV

QAAPQGVOLFPCNIEFQAGWKQGRPISSNRYQLQADGSLIIRLRPEAGIYS

CGSRHPQEPQEIQLRVTDGMALPEGQPHRPNPDLGHGPHRGTGABAGIYS

VLSSHPRPATRLRLDRTOPVVDASPGORILTCRAEGEPVPTIELWRDQOLVSPR

HQVQDGLSVISRDVDEGGYSCVAFNGDQRQWQVLRVLRILITGLPPAVTAE

DTARLLCVAGSEVNIWRNGLPIQADGHRVQSPDGLLIHNRUPDEGSGITCSA

FRGSQAVSRSTEYKVALPAPAAOSRDIGDKCIDQPELANCALILQALQGLGNEYSFC

CASCRFPQNAQPVWQOQ"

BASE COUNT 885 a 1276 c 1310 g 800 t

ORIGIN

Query Match 6.2%; Score 163.4; DB 10; Length 4271;

Best Local Similarity 50.1%; Pred. No. 5.4e-24;

Matches 496; Conservative 0; Mismatches 481; Indels 14; Gaps 3;

QY 1118 cctgcagcaagcgcctcgcctcgcctgagcagcagcagcagcagcagcagcagcag 1177

DB 283 CCTGCCACAGGAGAGCTCCAGACGGCGTGGGGATTTCGGGGCCGAGGAGTGGAG 342

QY 1178 cctttaactccagggaattcagcagcagcagcagcagcagcagcagcagcagcagcag 1237

DB 343 AGTTCGACGGCACACCTCCAGGCTCGCGCTACCGTGGCTGCCCTACTACGAGGCC 402

QY 1238 agggctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1297

DB 403 -----CTAATAAGTGTAGCTGAACATGCCCAAGGCGAGAACTTCTACTACAAAC 456

QY 1298 acactgaaaggtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1357

DB 457 ACAAGGACCCCTGGTGGATGGGACACCTCTGTGACCTGGCCAGCGGACATCTGTGTGG 516

QY 1358 cggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1417

DB 517 ACGGAGTCTGCCAGTGTGGTGTGTGATCATCAAGTACTAGACTCAATCAAGCAGGAGGACA 576

QY	1418	gctgtggagctgtgtgggggtgatattctacctctgcgccttgttttggggaaacctcactg	1477
Db	577	AGTGTCTGCAGTGTGGGGGTGATGGCTATCTCTCTACCCGGTACAGGAACCTTTGATG	636
QY	1478	accgagggggcccttggcttatcagaagactcttggattccagcgggagccttgcggc	1537
Db	637	GCAATGATCTAGCAGAGGCTACAAACCAAACTTCATCATTCGGCTGGGGCAACTAGTA	696
QY	1538	tccagattcccagctccggcctagctccaaactacctggcaactctgtggcccttggggcc	1597
Db	697	TTCCGATTGAGGAAGCCCTGCCAGCAGGAATTTCTGGCGGTGAAGAGATCCGTGGCG	756
QY	1598	ggtccatcatcaatgggaactcggctgtgggtcccttggctcctacagcggcggcggga	1657
Db	757	AGTACTATCTCAATGGGCACTGGACATTGAGGCAGCCAGGCTCTCCGGTGGCCAGCA	816
QY	1658	ccgtcttttcgatataacccgtctccagcggaggagggaagggagagctctgtcggctg	1717
Db	817	CAGTCTCAGTACGAACGGGGTGTGGAGGGGGACCTGGTCTGTAGCGGCTCCAGGCAC	876
QY	1718	aaggcccacacccagcctgtgagtctctatgattcttcaggaggaaaaaccaggcg	1777
Db	877	GGGGCCCCACCTCAGAGCCCTGTGCATCGAGCTCCTGAGTCAGGATGCCAACCCCGGAG	936
QY	1778	ttttttcaagtatgtcatctcttccactcctccaactcttgaaacccccaccacccagagc	1837
Db	937	TACACTACGAGTACTACTGCCCGGAATGATCTCTGGCC----GAGGCTTCAGCTGGAGC	992
QY	1838	ccctgtccccagcttcagccggagattctgagggctggagcccccacttgcctcggcac	1897
Db	993	CATGGCTCTCTGGGAGACTGCAAGCCGTGAGTGTGGTGGAGTCAACAGTCCCGCTGGTA	1052
QY	1898	ccgcgccagccgggacccaggcaccctccagcgtccaggtcgggattccccagatgcccg	1957
Db	1053	TTCTGCACATTGACAACGAGGCCATTCCAGACCACATGTGCCAACACCAGCCTCGGCC	1112
QY	1958	cccgccccatccccagg----acaccttgggtctccagctcgggtactggaaacagatg	2013
Db	1113	ACCCACCGAGCTTCATGCAACACTCAACTTGTCCGAAGACCAAGCGCTGGAAGGTAGGC	1172
QY	2014	ggacactctgcattccagcgtccctcgggaaaggtgtcttggcgcccaatttctctctgc	2073
Db	1173	CCATGGACGCCCTGCTCGGTCTCCTCGGGGTGGAGTCCAGTCTCGTCTGTCTACTGC	1232
QY	2074	atctcccttgagtcggagaggaactgtagt	2104
Db	1233	ATATCTCAGATGGGACTGGTGGCCAGGAAG	1263

RESULT 11  
 AX319854  
 LOCUS  
 DEFINITION  
 SEQUENCE 18 from Patent WO0183782.  
 ACCESSION  
 AX319854  
 VERSION  
 AX319854.1 GI:17901444  
 KEYWORDS  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS  
 Plowman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and  
 Payne,V.  
 TITLE  
 Novel proteases  
 JOURNAL  
 Patent: WO 0183782-A 18 08-NOV-2001;  
 Sugen, Inc. (US)  
 FEATURES  
 Location/Qualifiers  
 1..2805  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 745 a 677 c 733 g 650 t  
 BASE COUNT  
 ORIGIN

Query Match	4.3%	Score 113.2	DB 6	Length 2805
Best Local Similarity	50.0%	Fred. No. 1.7e-13		
Matches 317	Conservative	0	Mismatches 308	Indels 9
Gaps				
QY 1159	cgggccctgcagtcgcgagccctttaactccacaggaattcatggccagctgtatcagtcgg	1218		
DB 1039	CGGCGCTCAACAGTGTGCAGATATAACAGCAAACTTTCCGTGGATGGTTCTTCCAGTGG	1098		
QY 1219	gagccctcaactgaagtcacagggctccacagcgtgtgaactgaactaacccgcgcgcggc	1278		
DB 1099	AAACCTTATACAAAGATGGAAGAGAGATCGATGCAAACTGTACTGCAAGCTTGAGAAC	1158		
QY 1279	tccgcctctatgcctgcacactgaaaggtccagagtgagccctgtgtcagccctgga	1338		
DB 1159	TTTGAATTTTTTTTGGCAATGTCGGCAAGTGAAGATGAACCTCCCTGCTCCCAAAAC	1218		
QY 1339	gcccctgacatctgtgtgctggagcgtgtctgagccccgcgcgtgtgatggatccttggc	1398		
DB 1219	AAAAATGATGTTGTATGTACGGGGTTTGTGAATGATGAGGATGTGATCATGAACCTAGGC	1278		
QY 1399	tctggcagcgcctctgatggctgtggagctgtgtggggtgatgattctaccctgtgcctt	1458		
DB 1279	TCTAAGCAGTTTCAGATGCTTGTGGCGTTTGCRAAGGTGATAATTCAACTTGCAAGTTT	1338		
QY 1459	gttcggggaaacctcaactgacgcgagggggccccctggcctatcagaagatcttgtggatt	1518		
DB 1339	TATAAGGCCCTGTACCTCAACACAGCATAAAGCAAAATGAATATTATCCGGTGGTCTCAT	1398		
QY 1519	ccagcgggagccttgcgscctccagattccagcctccgccttagctccaaactaacctggca	1578		
DB 1399	CCAGCTGGCGCCGNAAGCATCGAATCCAGAGCTGCAGGTTTCCTCCAGTTACCTCGCA	1458		
QY 1579	cttcgtggccctggggggccgttccatcatcaatgggaaactggcgtgtgtgacccccctgg	1638		
DB 1459	GTTCCGAAGCCTCAGTCAAAAGATTTACCTACCGGGGGCTGGAGATCGACTGGCCTGGG	1518		
QY 1639	tcctacagggccggcggagcctcttcgtataaccctctcccccagggagggagggcaaa	1698		
DB 1519	GAGTTCCTTCGCTGGGACACGTTTGAATACACGCTCTTTCAA-----CCGC	1569		
QY 1699	ggggagctgtcgtcgtgaaagggccccaccacccagcctgtgtgagctctatgatcttt	1758		
DB 1579	CCGGAAGCTCTGTACGCGCCAGGGGCCCAAAATGAGAGCTGTCTTTGAAATTCGTATG	1629		
QY 1759	caggaggaaaacccagcggcttttttatcagtatg	1792		
DB 1639	CAAGCAAAAATCCAGGATACCTTGGAGATG	1663		
RESULT 12				
AX319852		3675 bp	DNA	linear
LOCUS	AX319852			
DEFINITION	Sequence 16 from Patent WO0183782.			
ACCESSION	AX319852			
VERSION	AX319852.1	GI:17901442		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
TITLE	1 (sites)			
JOURNAL	Florman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and			
FEATURES	Payne,V.			
source	Novel proteases			
	Patent: WO 0183782-A 16 08-NOV-2001;			
	Sugen, Inc. (US)			
	Location/Qualifiers			
	1..3675			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
BASE COUNT	874 a 1009 c 1019 g 773 t			

ORIGIN

Query Match 4.1%; Score 106.8; DB 6; Length 3675;  
Best Local Similarity 48.5%; Pred. No. 3.5e-12;  
Matches 331; Conservative 0; Mismatches 342; Indels 9; Gaps 1;

QY 1119 ctgcagcgaagcgcctgccccctgagcagccagccccccgccccctgagtcgagc 1178  
DB 1902 CTGCAACAGTCAGAAATGTCCCCGGACAGTGTGACTTCCTCGTGTGCTGACGTGCCGA 1961  
QY 1179 ctttaactccaggaattcatggccagctgtatcagtgaggagccctcactgaagtcga 1238  
DB 1962 GCACACAGCAGCAGATTTCAGAGGGGGGACACTACACAGTGGAGCCCTTACACTCAGTAGA 2021  
QY 1239 gggctccagcgcgtgtgaactgaactgcggcccccggtggcttcctccgttatgtccgtca 1298  
DB 2022 AGATCAGGACTTATGCAAACTCTACTGTATCGCAGAGGATTGATTCTTCTTCTTT 2081  
QY 1299 cactgaaggtccagatggagaccctgtgcagcctgagccctgcacatctgtgtggc 1358  
DB 2082 GTCAATAAAGTCAAGATGGGACTCCATGCTCGGAGGATAGCCGTAATGTTTGTATAGA 2141  
QY 1359 tggacgcgtctgagccccgctgtgatggatccttggctctgtggcagcgtccctgatgg 1418  
DB 2142 TGGGATATGTGAGAGAGTTGGATGTGACAAATGTCCTTGGATCTGATGCTGTTGAAGACGT 2201  
QY 1419 ctgtgagctctgtgggtgagtgattctacctctgctgcctgtgttcggggaacctactga 1478  
DB 2202 CTGTGGGTGTGTAAAGGGAATAAATCAAGCTTCACAGGATTCACAGGGGTCTCTACACCAA 2261  
QY 1479 ccgaggggccccctggctcagagatcttggatccagcggggagccttggcgtt 1538  
DB 2262 GCACCACCACCAACAGTATTATACATGGTCACCATCTCTCTGAGCCCGGAGTAT 2321  
QY 1539 ccagatggccagctccgctagctccactactctggcactctgtggccccctggggcccg 1598  
DB 2322 CCGACTTATGAATGAACGTCTCTACCTCTCTACATTTCTGTGCGCAATGCCCCCTCAGAA 2381  
QY 1599 gtccatcataatgggaactggcgtgtggtacccctggctgtctacagggccgggggac 1658  
DB 2382 GTACTACCTGAATGGGCACCTGGACCTGGACCTGGCCCGCCGGTACAAATTTTCGGGCAC 2441  
QY 1659 cgtcttcgatataacgctctccagggaggggaggggcaaaaggggagagtcctgtcggctga 1718  
DB 2442 TACTTTTCGACTACAGACGCTCTATATAG-----CCCGAGAACTTATCGCTAC 2492  
QY 1719 agggccaccaccagcctgtggtatgtctatgatcttccagggaggaacccagggcgt 1778  
DB 2493 TGGACCAACCAACAGACACTGATCTGGAGCTGCTGTTTCAGGGAAGAACCCGGGTGT 2552  
QY 1779 tttttatcagtatgtcatctct 1800  
DB 2553 TGCCTGGGAATACCTCATGCT 2574

RESULT 13  
AX342638  
LOCUS AX342638 2410 bp DNA linear PAT 12-JAN-2002  
DEFINITION Sequence 35 from Patent WO0198468.  
ACCESSION AX342638  
VERSION AX342638.1 GI:18152035

KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Yue.H., Elliott.V.S., Gandhi.A.R., Lal.P., Au-Young.J.,  
Tribouley.C.M., Deleageane.A.M., Baughn.M.R., Nguyen.D.B., Lee.E.A.,  
Hafalia.A., Khan.F.A., Wallia.N.K., Yao.M.G., Lu.D.A., Patterson.C.,  
Tang.Y.T., Walsh.R.T., Azimzal.Y., Ramkumar.J., Xu.Y. and Reddy.R.  
Patent: WO 0198468-A 35 27-DEC-2001;

FEATURES

Location/Qualifiers  
1..2410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 3750004Cbl"  
BASE COUNT 439 a 832 c 651 g 488 t  
ORIGIN

Query Match 3.8%; Score 99; DB 6; Length 2410;  
Best Local Similarity 48.2%; Pred. No. 1.6e-10;  
Matches 384; Conservative 0; Mismatches 395; Indels 18; Gaps 3;

QY 1134 ctgccccctgagcagccagccccccgccccctgagtcgagcagcctttaaactccagga 1193  
DB 310 CTGCCCCCAGGGGCTGTGCCCTTCAGAGACTTACAGTGTGCTGTACATAGCCGCC 369  
QY 1194 attcatggccagctgtatcagtgaggcccttcaatgaagtcagggtccccagcgtg 1253  
DB 370 TGTCTTGGGACCCAGAGACCTACCAGTGGTGGCTTCCATGGGCGCCCAACAGTG 429  
QY 1254 tgaactgaactgcggccccctggcttcctctctatgtccctcaactcaactgaagtcga 1313  
DB 430 CGACCTCAACTGCCTGGCTGAGGGGACGCTTCTA---CCACAGCTTCGGCCGCTCT 486  
QY 1314 ggaatggacctgtgcagcctgagccccctgacatctgtgtgctgagcgtgtctgag 1373  
DB 487 GCAGCGACCCCTGCACCCCGGGTGCACAGGGGTGTGCTGTGCTGCGCTGCCCTAG 546  
QY 1374 cccggctgtgatggatccttggctctggcagcgtcctgatggtcgtgagtcgtg 1433  
DB 547 CCGCGCTGTGTATGGTGTGGGCTCGGTCGCTCGAGACCGCTGTGCGCTCGG 606  
QY 1434 ggggatattctacctgtcgccctgtttcggggaacctcaactcaactgaggggccccct 1493  
DB 607 AGGCCCAACAGACTCGTCCCTTTGCTGCAGCGGTGTTCTGACGCCGTGCTTCGC 666  
QY 1494 gggctatcagaagatcctgtgattccagcggagccttgcgctccagattgcccagct 1553  
DB 667 TGGGTACTGGAACGTGACCTGTATCCCGAGGGGCCACACATCCGCGTGAACACAG 726  
QY 1554 ccggcctagctcc-----aactacctggacttcgtggccctggggcggtccatcat 1607  
DB 727 GAGCCGCAACCACTGGGTATCTTAGATCACTATGATGGGGGCGATGGGCGCTACG 786  
QY 1608 caatgggaactgggctgtggtatcccccctggctcctacagggccgagcagcgtcttctc 1667  
DB 787 TAATGGGCACTGGGTGTGTCAGCCACACAGGACCTACGAGCGCGCGCATGTGGT 846  
QY 1668 atataacgctcctccagggaggggcaaaaggggagagtcgtcggctgaagccccac 1727  
DB 847 CTACA-----CCCGAGACACAGGGCCCCCAGGAGACATTGCAAGCAGCGCGGCCAC 897  
QY 1728 caccagcctgtggtatgtctatgatcttccagggaggaacccagcgttttttatca 1787  
DB 898 CTCCTCATGCTCTCTACAGGTCTCTCTGACAGAGCCCAACCTGGCATCGAGTTGA 957  
QY 1788 gtagtcatctcttcacccctccatcttccttgaagccccccacccagccccctgtccc 1847  
DB 958 GTTCTGGCTCCCTCGGGAGCGCTACAGCCCTTCAGGCTCGTGTGAGGCCCTGGGCTG 1017  
QY 1848 ccagcttcagcggagattctgaggtgagccccccactgtcggcaccgcccccagc 1907  
DB 1018 GCCCTGAGGAGCCTCAGCCCCGGGGGTGGAGCCTCAGCCCCCGGAGGCCCTGTGT 1077

RESULT 14  
AC014412



LOCUS AC014412 86446 bp DNA linear HTG 16-NOV-1999  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
 ACCESSION AC014412  
 VERSION AC014412.1 GI:6436923  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 86446)  
 AUTHORS Adams, M. and Venter, J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
 COMMENT This sequence was identified as CDM:10212722 by the submitter.  
 \* NOTE: This is a 'working draft' sequence.  
 \* this sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES  
 source  
 1. 86446  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 BASE COUNT 23099 a 20123 c 20064 g 23160 t  
 ORIGIN  
 Query Match 3.7%; Score 97; DB 2; Length 86446;  
 Best Local Similarity 47.6%; Pred. No. 2.3e-10;  
 Matches 323; Conservative 0; Mismatches 350; Indels 6; Gaps 1;  
 QY 1118 cctgcagccaagcgcctcgcctccctcagcagccagcccccggcctgcgtgcgagc 1177  
 DB 8870 CTTGCAACACCCAGACTGTCGGAGGAGAAATCCGACTTCGGGGCGAGAGTGTC 8929  
 QY 1178 ctttaactccaggaattcattggtcagcgtatcagtgaggagcccttcgaagtc 1237  
 DB 8930 GCTTCGACGAGCAGTGTGAGTCTTACAGTGGGTGGCGGTACACAGATGCAC 8989  
 QY 1238 aggggtccagcgcgtgtgaactgaactgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1297  
 DB 8990 -----CCAAATCCTCGAGCTGAATGTATGCGGAGGCGGAGCGGTCTACTACCG 9043  
 QY 1298 acactgaaaggtccagagatggagccctgttcagcctggagccctgcacatctgtg 1357  
 DB 9044 AGCGGAGAGGTGGTGGAGCGGAAACCGCTGCAACGACGACCTCGAGTGGCTAA 9103  
 QY 1358 ctggagcgtgtgtgagcccccgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1417  
 DB 9104 ACGGGGAGTGCATCGCGGTGGTGGGACATGATGCTGGGTAGCGATGCCAAGGAG 9163  
 QY 1418 gctgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1477  
 DB 9164 AGTCCGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9223  
 QY 1478 accagggggcccccctggggtatcagaagatctgtgtgtgtgtgtgtgtgtgtgtgt 1537  
 DB 9224 CGAAGGATCGGCTCGGATACATGACTTGTCTGCTCTCCAGAGGGAGCCACCA 9283  
 QY 1538 tcagattgcccagctccggtcagctcagctcagctcagctcagctcagctcagctcag 1597  
 DB 9284 TTGCTATCGAGGAGACTGTCCCTTCAGCAATTACCTGGCTCCCGGAAACACAGTGG 9343  
 QY 1598 ggtccatcatcaatgggaactgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1657  
 DB 9344 ACTACTACCTGAACGGGAGTGGCGATCGACTTCCACGCCCCATGTTCTTCGCGA 9403  
 QY 1658 cgtcttcgatataaccgctcccccagggagagggagggagggagggagggagggaggg 1717  
 DB 9404 CGTGGTGAAGTACTCAACAGGAAGCAATGGGCTTCGCCGCTCCCGCAACCACTG 9463

QY 1718 aagggcccaacccagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1777  
 DB 9464 CGGTCCTCATCTCGAGAGCCTCTTCTATGTTATGTTGGTGGAGAGAGACATCAGCC 9523  
 QY 1778 tttttatcagatgtcat 1796  
 DB 9524 TCGACTACGATACAGCAT 9542  
 RESULT 15  
 AC009213 182525 bp DNA linear INV 17-FEB-2001  
 LOCUS Drosophila melanogaster, chromosome 3R, region 98D-98D, BAC clone  
 DEFINITION BACR09F18, complete sequence.  
 ACCESSION AC009213  
 VERSION AC009213.5 GI:12957654  
 KEYWORDS HTG.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 182525)  
 AUTHORS Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Ananides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busan, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferrier, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Shapiro, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
 TITLE Sequencing of Drosophila chromosome 3R, region 98D-98D  
 JOURNAL Unpublished  
 2 (bases 1 to 182525)  
 AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.K., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snit, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 COMMENT On Feb 17, 2001 this sequence version replaced gi:6957913.  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgpf@fruitfly.berkeley.edu](mailto:bdgpf@fruitfly.berkeley.edu).  
 FEATURES  
 source  
 1. 182525  
 /organism="Drosophila melanogaster"  
 /strain="y: cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 /map="98D-98D"  
 /clone="BACR09F18 (D812)"  
 /clone\_lib="RBCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in

BASE COUNT 49013 a 41841 c 41897 g 49774 t  
ORIGIN pBACE3.6)"

Query Match 3.7%; Score 97; DB 3; Length 182525;  
Best Local Similarity 47.6%; Pred. No. 2e-10;  
Matches 323; Conservative 0; Mismatches 350; Indels 6; Gaps 1;

QY 1118 cctgagcccaagccctccccctgagcagcagccccccggccctgcaagtgcgag 1177  
|||||  
Db 70744 CTTGCAACACCCAAAGACTCTCCGGAGGAAGATCCGACTTCGGGGCGCAGAGTGCTCCC 70685  
QY 1178 cctttaactcccaggaaatcatgggccaagtgtatcatagtgaggagcccttcaactgaagtcc 1237  
|||||  
Db 70684 GCTTCGACAGGAGCAGATTGATGGAGTCTTACGAGTGGGTGCGGTACACGAATGCAC 70625  
QY 1238 agggctccagcagcgtgtgaactgaactgcggcccccgtggtccgcttctatgtccgtc 1297  
|||||  
Db 70624 -----CCAACTCCCTCGAGCTGAATGTATGCCAAAGGGCGAGCGGTCTACTACCGTTC 70571  
QY 1298 acactgaaaggtccaggaatggagccctgtgcagcgtggagccctgacatctgtgtgg 1357  
|||||  
Db 70570 AGCGGAGAGGTGGTGGAGACCCGCTGCACGACAAAGACCTCGACGTTGCGTAA 70511  
QY 1358 ctggagcgtgtgtgagccccggcgtgtgattggtgagatccctgtgctctggaaggcctcctgatg 1417  
|||||  
Db 70510 ACGGGAGTGCATCGCGTGGCTGCGACATGCTGGGTAGCGATGCCAAGGAGGACA 70451  
QY 1418 gctgtgagctgtgtgggtgtgattgattacctgtgcctgtgttcggggaaacctcactg 1477  
|||||  
Db 70450 AGTGCCGAAGTGTGTGTGTGATGGCAGACCTGCNAGACCAATTCGACACCAATCACCA 70391  
QY 1478 accgagggggccccctggcctatcagaagatttgtggtattccagogggagccttgcggc 1537  
|||||  
Db 70390 CGAAGGATCTGGTCTCTGGATACAATGACTTGTGTCTCTCCAGAGGGAGCCACTAACA 70331  
QY 1538 tccagattcccaagctccggcctagctccaactacctgacacttcgtggccctggggcc 1597  
|||||  
Db 70330 TTGCTATCGAGGAGACTGTGCCCTCCAGCAATTACCTGGCCCTGCCGGAACCAACAGTGGAC 70271  
QY 1598 ggtccatcatcaatgggaactgggctgtggtatccccctgggtccctacagggccggcgga 1657  
|||||  
Db 70270 ACTACTACCTGAACGGCGATTGGCGCATCGACTTCACGCCCCCATGTTCTTCGGGAAC 70211  
QY 1658 cegtctttcgataaccgtccctccagggagggagggcaaggagagagctgtcggtg 1717  
|||||  
Db 70210 CGTGGTGGAACTACCAACGAAGGCCAATGGGTTCGCCGCTCCCGCAACCACTGACCTGCA 70151  
QY 1718 aagccccaccaccagcctgtggtgtctatatgatctttcaggaggaaacccaggcg 1777  
|||||  
Db 70150 GCGTCCCATCTCGGAGAGCCCTTTCATAGTTATGTTGTGCGAGGAGAGAACAATCAGCC 70091  
QY 1778 tttttatcagtatgtcat 1796  
|||||  
Db 70090 TCGACTACGAGTACGAT 70072

Search completed: July 23, 2002, 17:35:57  
Job time: 16247 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 18:27:20 ; Search time 292.12 Seconds  
(without alignments)  
15481.147 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atgaggaactggactggcag.....ccctagcctttccagcatag 2634

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1259	47.8	2680	22 AAC91178	ADAM gene #2. Hom
2	377	14.3	1020	22 AAD16773	Human novel protei
3	377	14.3	1998	22 AAD16758	Human novel protei
4	54	2.1	320	21 AAC16560	Human secreted pro
5	23	0.9	704	22 AAH93398	Human protein enco
6	23	0.9	814	22 AAH08271	Human CDNA clone (
7	23	0.9	2964	22 AAH16636	Human CDNA sequenc
8	20	0.8	693	23 AAS82152	DNA encoding novel
9	20	0.8	1827	23 AAS66958	DNA encoding novel

10	20	0.8	1927	22 AAH76376	Murine adipocytes-
11	20	0.8	1949	22 AAC84391	Mouse A236 polyep
12	20	0.8	1949	22 AAC84416	Mouse A236 variant
13	20	0.8	1949	22 AAC84417	Mouse A236 variant
14	20	0.8	1949	22 AAC84418	Mouse A236 variant
15	20	0.8	1995	22 AAF63462	Oligonucleotide #2
16	20	0.8	2013	23 AAS83442	DNA encoding novel
17	20	0.8	2153	23 AAS78212	DNA encoding novel
18	20	0.8	17069	22 AAK80906	Human immune/haema
19	20	0.8	17069	22 AAK83892	Human immune/haema
20	20	0.8	32206	22 AAK89374	Human digestive sy
21	19	0.7	308	22 AAS28913	Human immunoglobul
22	19	0.7	410	18 AAT51050	Glycoprotein gp105
23	19	0.7	410	21 AAAT70394	Partial gp105 prot
24	19	0.7	609	22 AAL22948	Human breast cance
25	19	0.7	647	22 AAH06718	Human CDNA clone (
26	19	0.7	669	22 AAH16868	Human digestive sy
27	19	0.7	671	22 AAK88735	Human CDNA encodin
28	19	0.7	671	22 AAS31769	Human liver associ
29	19	0.7	724	22 AAS26853	Human CDNA encodin
30	19	0.7	1790	22 AAL05360	Human reproductive
31	19	0.7	2055	19 AAV88524	Nucleotide sequenc
32	19	0.7	2055	21 AA260905	DNA encoding a hum
33	19	0.7	2058	21 AA291643	Human notch ligand
34	19	0.7	2159	22 AAS21287	Human CDNA sequenc
35	19	0.7	2184	21 AAZ60918	A human delta-rel
36	19	0.7	2437	22 AAH18557	Human CDNA sequenc
37	19	0.7	2442	22 ABA06484	Human CDNA SEQ ID
38	19	0.7	2508	22 AAH14855	Human CDNA sequenc
39	19	0.7	2615	22 AAH18513	Human CDNA sequenc
40	19	0.7	2670	23 ABL04247	Drosophila melanog
41	19	0.7	2800	19 AAV68523	Nucleotide sequenc
42	19	0.7	2897	24 AAD26454	Human kinase PKIN-
43	19	0.7	3339	20 AAX16297	Human delta-2 prot
44	19	0.7	4426	23 ABL17454	Drosophila melanog
45	19	0.7	4878	23 ABL04246	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAC91178  
ID AAC91178 standard; DNA; 2680 BP.

XX AC AAC91178;  
XX DT 20-MAR-2001 (first entry)  
XX DE ADAM gene #2.  
XX KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;  
XX KW ovarian; ss.  
XX OS Homo sapiens.  
XX PN WO200073323-A2.  
XX PD 07-DEC-2000.  
XX PF 25-MAY-2000; 2000WO-US14308.  
XX PR 27-MAY-1999; 99US-0136388.  
XX PR 09-JUL-1999; 99US-0142930.  
XX PR 28-JAN-2000; 2000US-0178717.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;  
XX DR WPI; 2001-016507/02.  
XX PT Seven nucleic acid molecules encoding ADAM polypeptides containing a

PT disintegrin and metalloprotease domain, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
XX Claim 1; Page 266; 287pp; English.  
XX  
CC The present invention relates to seven members of the ADAMs (proteins  
CC which contain A Disintegrin And Metalloprotease domain) protein family.  
CC The ADAMs proteins and DNA may be used to treat disease, as a food  
CC additive or preservative, for chromosome identification, as probes  
CC for diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the  
CC gene therapy of breast and ovarian cancer.  
XX  
XX Sequence 2680 BP; 510 A; 864 C; 767 G; 537 T; 2 other;

Query Match 47.8%; Score 1259; DB 22; Length 2680;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 ctgaaaggtccaggatggagacctgtgcagcctggagccctgacatctgtgtgctg 1360  
DB 3 ctgaaaggtccaggatggagacctgtgcagcctggagccctgacatctgtgtgctg 62  
QY 1361 gacgtctgtgagcccgctgtgatggatccttgcttgagcgtctctgatgctt 1420  
DB 63 gacgtctgtgagcccgctgtgatggatccttgcttgagcgtctctgatgctt 122  
QY 1421 gtggagctgtgggggtgatgttacctgtgccttcttctgggaaacctcaactgacc 1480  
DB 123 gtggagctgtgggggtgatgttacctgtgccttcttctgggaaacctcaactgacc 182  
QY 1481 gagggggccccctgggtatcaagaatctgtgattccagcagcgtctgcgctcc 1540  
DB 183 gagggggccccctgggtatcaagaatctgtgattccagcagcgtctgcgctcc 242  
QY 1541 agatgcccagctccgctgagctcaactacctggaactctgctgcccctggggccggt 1600  
DB 243 agatgcccagctccgctgagctcaactacctggaactctgctgcccctggggccggt 302  
QY 1601 ccatcatcaatgggaactgggtgtggatccccctgggtccctacagggccgggagccg 1660  
DB 303 ccatcatcaatgggaactgggtgtggatccccctgggtccctacagggccgggagccg 362  
QY 1661 tcttcgataaaccttctccagggagaggccaaaggaggagctctgtcggctggaag 1720  
DB 363 tcttcgataaaccttctccagggagaggccaaaggaggagctctgtcggctggaag 422  
QY 1721 gcccccacacccagcctgtgagcttatatgatctttcaggaggaaaacccaggcgctt 1780  
DB 423 gcccccacacccagcctgtgagcttatatgatctttcaggaggaaaacccaggcgctt 482  
QY 1781 ttatcagtagtgcattctctccctcccaactccttgagaaacccacagagccccc 1840  
DB 483 ttatcagtagtgcattctctccctcccaactccttgagaaacccacagagagccccc 542  
QY 1841 ctgtccccagcttcagccggagattctgaggggtggagcccccaactgtctcgggacccc 1900  
DB 543 ctgtccccagcttcagccggagattctgaggggtggagcccccaactgtctcgggacccc 602  
QY 1901 gccacggccgagccacccagcactccagctcaggtcgagatcccccagatgcccgccc 1960  
DB 603 gccacggccgagccacccagcactccagctcaggtcgagatcccccagatgcccgccc 662  
QY 1961 cgccccatccagagacacccctgggtctctccagctcgtactggaaacagtgaggacact 2020  
DB 663 cgccccatccagagacacccctgggtctctccagctcgtactggaaacagtgaggacact 722  
QY 2021 ctgcagctcagcgtcctcgggaaagggtgtgtggcccccattttctctcgtcatctccc 2080  
DB 723 ctgcagctcagcgtcctcgggaaagggtgtgtggcccccattttctctcgtcatctccc 782

QY 2081 gtgagtcgggagagaaactggatgaacacagctgtgcgcgggtgcagggccccccagcct 2140  
DB 783 gtgagtcgggagagaaactggatgaacacagctgtgcgcgggtgcagggccccccagcct 842  
QY 2141 cccctgaacccctgcacagcgcaccccccatactgcccatactgggaggtggcagtggaacat 2200  
DB 843 cccctgaacccctgcacagcgcaccccccatactgcccatactgggaggtggcagtggaacat 902  
QY 2201 cctgcagcgcctcctgtgccccggcaccacagcagccagctgcagtgccggcagggaat 2260  
DB 903 cctgcagcgcctcctgtgccccggcaccacagcagccagctgcagtgccggcagggaat 962  
QY 2261 ttgggggggggtggctcctcctggtgccccgggagcgtgtggacatctccccgggcccaca 2320  
DB 963 ttgggggggggtggctcctcctggtgccccgggagcgtgtggacatctccccgggcccaca 1022  
QY 2321 tcacccagctcttgccagctgcgcctctgtgcccattgggaagtggctctcttggagcc 2380  
DB 1023 tcacccagctcttgccagctgcgcctctgtgcccattgggaagtggctctcttggagcc 1082  
QY 2381 agtgcctcgtcggtgccccgggcccagagaagccggcaggttgcgtgtgttgggaaca 2440  
DB 1083 agtgcctcgtcggtgccccgggcccagagaagccggcaggttgcgtgtgttgggaaca 1142  
QY 2441 acggtgatgaagtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2500  
DB 1143 acggtgatgaagtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1202  
QY 2501 aggcctgtgacatggggccctgtactactgctgttccacagcagctggagctcccaag 2559  
DB 1203 aggcctgtgacatggggccctgtactactgctgttccacagcagctggagctcccaag 1261

RESULT 2

AAD16773  
ID AAD16773 standard; cDNA; 1020 BP.  
XX AC AAD16773;  
XX DT 22-NOV-2001 (first entry)  
XX DE Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:34.  
XX DE Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
KW autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;  
KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;  
KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
KW anti-sense therapy; endocrine disorder; leukaemia; ss.

XX OS Homo sapiens.  
XX PH Key Location/Qualifiers  
XX CDS 65..847  
XX FT /\*tag= a  
XX FT /product= "Human novel protein"  
XX FT /transl\_except= (pos:767..769, aa:Xaa)  
XX FT /note= "CDS does not include start codon; Xaa is an  
XX FT unknown amino acid"  
XX FT /partial  
XX PN WO200155202-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01325.  
XX PF 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.

PR	17-MAR-2000;	2000US-0190076.	PR	13-OCT-2000;	2000US-0239937.
PR	18-APR-2000;	2000US-0198123.	PR	20-OCT-2000;	2000US-0240960.
PR	19-MAY-2000;	2000US-0205515.	PR	20-OCT-2000;	2000US-0241221.
PR	20-JUN-2000;	2000US-0209467.	PR	20-OCT-2000;	2000US-0241785.
PR	28-JUN-2000;	2000US-0214886.	PR	20-OCT-2000;	2000US-0241786.
PR	30-JUN-2000;	2000US-0215135.	PR	20-OCT-2000;	2000US-0241787.
PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000;	2000US-0216680.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	14-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0245474.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0245475.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224518.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0248523.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0248524.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0248525.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0248526.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0248527.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0248528.
PR	14-AUG-2000;	2000US-0225271.	PR	08-NOV-2000;	2000US-0248532.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0248609.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0248610.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0248611.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0248613.
PR	18-AUG-2000;	2000US-0226279.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226681.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0226868.	PR	17-NOV-2000;	2000US-0249209.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249210.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249211.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249215.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249216.
PR	03-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249217.
PR	03-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249244.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232080.	PR	01-DEC-2000;	2000US-0250160.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250391.
PR	12-SEP-2000;	2000US-0231968.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0256179.

Isolated polypeptide for treating, preventing and/or proposing medical disorders and also for testing and detection e.g. diagnosis and screening for agonists -

Claim 1; SEQ ID No 34; 469pp; English.

AAD16750-AAD16775 represent cDNAs corresponding to novel human protein



PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX	(HUMA-)	HUMAN GENOME SCI INC.
PR	26-SEP-2000;	2000US-0235484.	PA		
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	DR	WPI; 2001-451925/48.	
PR	29-SEP-2000;	2000US-0236367.	DR	P-PSDB; AAE09698.	
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	PT	Isolated polypeptide for treating, preventing and/or prognosing	
PR	29-SEP-2000;	2000US-0236370.	PT	medical disorders and also for testing and detection e.g. diagnosis and	
PR	02-OCT-2000;	2000US-0236802.	PT	screening for agonists -	
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237038.	XX		
PR	02-OCT-2000;	2000US-0237039.	PS	Claim 1; SEQ ID No 19; 469pp; English.	
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	CC	AAD16750-AAD16775 represent cDNAs corresponding to novel human protein	
PR	13-OCT-2000;	2000US-0239937.	CC	genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777	
PR	13-OCT-2000;	2000US-0239937.	CC	-AAD16780 represent novel human genomic DNA fragments. The novel proteins	
PR	20-OCT-2000;	2000US-0240960.	CC	and their DNAs are useful for diagnosing, treating, preventing and/or	
PR	20-OCT-2000;	2000US-0241221.	CC	prognosing inflammatory disorders (bursitis or tendonitis); neural	
PR	20-OCT-2000;	2000US-0241785.	CC	disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system	
PR	20-OCT-2000;	2000US-0241786.	CC	disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);	
PR	20-OCT-2000;	2000US-0241787.	CC	muscular disorders; reproductive disease; gastrointestinal disorders	
PR	20-OCT-2000;	2000US-0241808.	CC	(malabsorption syndrome, Crohn's disease); pulmonary disorders;	
PR	20-OCT-2000;	2000US-0241809.	CC	cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);	
PR	20-OCT-2000;	2000US-0241826.	CC	renal disorders (glomerulonephritis, nephrotic syndrome); cancerous	
PR	01-NOV-2000;	2000US-024617.	CC	disease and conditions (breast cancer); hyperproliferative disorders	
PR	08-NOV-2000;	2000US-0246474.	CC	(leukaemia, hyperplasia); tumours; foetal and developmental	
PR	08-NOV-2000;	2000US-0246475.	CC	abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,	
PR	08-NOV-2000;	2000US-0246476.	CC	asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine	
PR	08-NOV-2000;	2000US-0246477.	CC	disorders; pregnancy-related disorders and infections. The novel protein	
PR	08-NOV-2000;	2000US-0246478.	CC	DNA is useful in gene therapy and anti-sense therapy. The proteins can	
PR	08-NOV-2000;	2000US-0246523.	CC	also be used to aid wound healing and epithelial cell proliferation,	
PR	08-NOV-2000;	2000US-0246524.	CC	to prevent skin aging due to sunburn, to maintain organs before	
PR	08-NOV-2000;	2000US-0246525.	CC		
PR	08-NOV-2000;	2000US-0246526.			
PR	08-NOV-2000;	2000US-0246527.			
PR	08-NOV-2000;	2000US-0246528.			
PR	08-NOV-2000;	2000US-0246528.			
PR	08-NOV-2000;	2000US-0246532.			
PR	08-NOV-2000;	2000US-0246609.			
PR	08-NOV-2000;	2000US-0246610.			
PR	08-NOV-2000;	2000US-0246611.			
PR	08-NOV-2000;	2000US-0246613.			
PR	17-NOV-2000;	2000US-0249207.			
PR	17-NOV-2000;	2000US-0249208.			
PR	17-NOV-2000;	2000US-0249209.			
PR	17-NOV-2000;	2000US-0249210.			
PR	17-NOV-2000;	2000US-0249211.			
PR	17-NOV-2000;	2000US-0249212.			
PR	17-NOV-2000;	2000US-0249213.			
PR	17-NOV-2000;	2000US-0249214.			
PR	17-NOV-2000;	2000US-0249215.			
PR	17-NOV-2000;	2000US-0249216.			
PR	17-NOV-2000;	2000US-0249217.			
PR	17-NOV-2000;	2000US-0249218.			
PR	17-NOV-2000;	2000US-0249219.			
PR	17-NOV-2000;	2000US-0249244.			
PR	17-NOV-2000;	2000US-0249245.			
PR	17-NOV-2000;	2000US-0249264.			
PR	17-NOV-2000;	2000US-0249265.			
PR	17-NOV-2000;	2000US-0249267.			
PR	17-NOV-2000;	2000US-0249297.			
PR	17-NOV-2000;	2000US-0249299.			
PR	17-NOV-2000;	2000US-0249300.			
PR	01-DEC-2000;	2000US-0250160.			
PR	01-DEC-2000;	2000US-0250391.			
PR	05-DEC-2000;	2000US-0251030.			
PR	05-DEC-2000;	2000US-0251988.			
PR	05-DEC-2000;	2000US-0256719.			
PR	06-DEC-2000;	2000US-0251479.			
PR	08-DEC-2000;	2000US-0251856.			
PR	08-DEC-2000;	2000US-0251868.			
PR	08-DEC-2000;	2000US-0251869.			

Query Match 14.3%; Score 377; DB 22; Length 1998;  
Best Local Similarity 99.8%; Pred. No. 1,7e-160;  
Matches 427; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	2132	ccccagctccctgaacccctgccacggcaccoccatgcccccactactggggagctggcg	2191
DB	81	ccccagctccctgaacccctgccacggcaccoccatgcccccactactggggagctggcg	140
QY	2192	agtggacatcctgcagcgcctcctgtgcccgcgaccacccagcagctgcagtgcc	2251
DB	141	agtggacatcctgcagcgcctcctgtgcccgcgaccacccagcagctgcagtgcc	200
QY	2252	ggcaggaaattgggggggtggctcctcgtgcccgcgagcgtgtggacatctcccc	2311
DB	201	ggcaggaaattgggggggtggctcctcgtgcccgcgagcgtgtggacatctcccc	260
QY	2312	ggcccaacatcaccagctgtgccagctgcgcctctgtggccattgggaagttggtctc	2371
DB	261	ggcccaacatcaccagctgtgccagctgcgcctctgtggccattgggaagttggtctc	320
QY	2372	cttggagccagtgctccgtgcggtgcccgcgggcccagagaaagccggcaggttcgctgtg	2431
DB	321	cttggagccagtgctccgtgcggtgcccgcgggcccagagaaagccggcaggttcgctgtg	380
QY	2432	tggggaacacgctgatgaagtgcagcagcagagagtgctgcagcagcccccacagcccc	2491
DB	381	tggggaacacgctgatgaagtgcagcagcagagagtgctgcagcagcccccacagcccc	440
QY	2492	ccagcagagagcctgtgacatgggcccctgtactactcctcgttccacagagctgga	2551
DB	441	ccagcagagagcctgtgacatgggcccctgtactactcctcgttccacagagctgga	500
QY	2552	gctccaaag	2559
DB	501	gctccaaag	508



RESULT 4  
AAC16560 standard; cDNA; 320 BP.  
XX  
AC AAC16560;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 20635.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EF1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1; SEQ ID 20635; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
SQ Sequence 320 BP; 50 A; 97 C; 117 G; 53 T; 3 other;

Query Match 2.1%; Score 54; DB 21; Length 320;  
Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagaaactgactggcagccctgctgtatctgctgctgtctgtcttc 54  
|||||  
Db 267 atgagaaactgactggcagccctgctgtatctgctgctgtctgtcttc 320  
|||||

RESULT 5  
AAH99398  
ID AAH99398 standard; cDNA; 704 BP.  
XX  
AC AAH99398;  
XX  
DT 16-OCT-2001 (first entry)  
XX  
DE Human protein encoding cDNA sequence SEQ ID NO:233.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnery; antileuk; antidiabetic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153455-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 22-DEC-2000; 2000WO-US35017.  
XX  
PR 23-DEC-1999; 99US-0471275.  
XX  
PR 21-JAN-2000; 2000US-048725.  
XX  
PR 25-APR-2000; 2000US-0552317.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-457603/49.  
XX  
DR P-PSDB; AAM25457.  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
PT Claim 1; Page 401-402; 1217pp; English.  
XX  
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
CC AAM25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
CC antileuk; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
SQ Sequence 704 BP; 184 A; 158 C; 195 G; 167 T; 0 other;

Query Match 0.9%; Score 23; DB 22; Length 704;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1600 tccatcatcaatgggaactgggc 1622  
|||||  
Db 445 tccatcatcaatgggaactgggc 467

RESULT 6  
AAH08271 ID AAH08271 standard; cDNA; 814 BP.  
XX AC AAH08271;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA clone (5'-primer) SEQ ID NO:5106.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-0116126.  
XX PR 29-JUL-1999; 99JP-0248036.  
XX PR 27-AUG-1999; 99JP-0300253.  
XX PR 11-JAN-2000; 2000JP-0118776.  
XX PR 02-MAY-2000; 2000JP-0183767.  
XX PR 09-JUN-2000; 2000JP-0241899.  
XX PA (HELI-) HELIX RES INST.  
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WP1: 2001-318749/34.  
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602  
XX PT full-length cDNAs defined in the specification, and for the detection  
XX PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX PT full-length cDNAs -  
XX PS Claim 1; SEQ ID 5106; 2537pp + CD ROM; English.  
XX CC The present invention describes primer sets for synthesizing 5602  
XX CC full-length cDNAs defined in the specification. Where a primer set  
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX CC to the complementary strand of a polynucleotide which comprises one of  
XX CC the 5602 nucleotide sequences defined in the specification, where the  
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX CC of an oligonucleotide comprising a sequence complementary to the  
XX CC complementary strand of a polynucleotide which comprises a 5'-end  
XX CC sequence and an oligonucleotide comprising a sequence complementary to a  
XX CC polynucleotide which comprises a 3'-end sequence, where the  
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of  
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in  
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX CC particularly full-length cDNAs. The primers are also useful for the  
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by  
XX CC the full-length cDNAs. The primers allow obtaining of the full-length  
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX CC represent oligonucleotides, all of which are used in the exemplification  
XX CC of the present invention.  
XX SQ Sequence 814 BP; 205 A; 235 C; 235 G; 136 T; 3 other;

Query Match 0.9%; Score 23; DB 22; Length 814;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1600 tccatcatcaatgggaactgggc 1622

Db 36 tccatcatcaatgggaactgggc 58  
RESULT 7  
AAH16636 ID AAH16636 standard; cDNA; 2964 BP.  
XX AC AAH16636;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA sequence SEQ ID NO:15752.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-0116126.  
XX PR 29-JUL-1999; 99JP-0248036.  
XX PR 27-AUG-1999; 99JP-0300253.  
XX PR 11-JAN-2000; 2000JP-0118776.  
XX PR 02-MAY-2000; 2000JP-0183767.  
XX PR 09-JUN-2000; 2000JP-0241899.  
XX PA (HELI-) HELIX RES INST.  
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WP1: 2001-318749/34.  
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602  
XX PT full-length cDNAs defined in the specification, and for the detection  
XX PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX PT full-length cDNAs -  
XX PS Claim 8; SEQ ID 15752; 2537pp + CD ROM; English.  
XX CC The present invention describes primer sets for synthesizing 5602  
XX CC full-length cDNAs defined in the specification. Where a primer set  
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX CC to the complementary strand of a polynucleotide which comprises one of  
XX CC the 5602 nucleotide sequences defined in the specification, where the  
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX CC of an oligonucleotide comprising a sequence complementary to the  
XX CC complementary strand of a polynucleotide which comprises a 5'-end  
XX CC sequence and an oligonucleotide comprising a sequence complementary to a  
XX CC polynucleotide which comprises a 3'-end sequence, where the  
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of  
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in  
XX CC the specification. The primer sets can be used in antisense therapy and  
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX CC particularly full-length cDNAs. The primers are also useful for the  
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by  
XX CC the full-length cDNAs. The primers allow obtaining of the full-length  
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX CC represent oligonucleotides, all of which are used in the exemplification  
XX CC of the present invention.  
XX SQ Sequence 2964 BP; 705 A; 862 C; 796 G; 601 T; 0 other;

Query Match 0.9%; Score 23; DB 22; Length 2964;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1500 tccatcatcaatgggaactgggc 1622  
|||||  
Db 36 tccatcatcaatgggaactgggc 58

RESULT 8  
ID AAS82152/c  
ID AAS82152 standard; cDNA; 693 BP.  
AC AAS82152;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17956.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
PI  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG17965.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 17956; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 693 BP; 159 A; 188 C; 204 G; 142 T; 0 other;

Query Match 0.8%; Score 20; DB 23; Length 693;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2229 ccagcaccgccagctgcagt 2248  
|||||  
Db 75 CCAGCACCGCCAGCTGCAGT 56  
RESULT 9  
ID AAS66958  
ID AAS66958 standard; cDNA; 1827 BP.  
AC AAS66958;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #2762.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
PI  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG02771.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 2762; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1827 BP; 374 A; 528 C; 512 G; 413 T; 0 other;

Query Match 0.8%; Score 20; DB 23; Length 1827;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2229 ccagcaccgccagctgcagt 2248

Db 762 ccagcaccgcccagctgcagt 781  
|||||  
RESULT 10  
AAH76376  
ID AAH76376 standard; DNA; 1927 BP.  
XX  
AC AAH76376;  
XX  
DT 30-NOV-2001 (first entry)  
XX  
DE Murine adipocytes-derived protein encoding DNA.  
XX  
KW Adipocyte; murine; differentiation; obesity; hyperlipemia; diabetes;  
KW atherosclerosis; ds.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT CDS 289..1410  
FT /\*tag= a  
XX  
PN WO200166720-A1.  
XX  
PD 13-SEP-2001.  
XX  
PF 09-MAR-2001; 2001WO-JP01863.  
XX  
PR 10-MAR-2000; 2000JP-0072502.  
XX  
PA (KITA/) KITAMURA T.  
PA (TSUR/) TSURUGA H.  
XX  
PI Kitamura T, Tsuruga H;  
XX  
DR WPI; 2001-565585/63.  
DR P-PSDB; AAB85862.  
XX  
XX Genes associated with adipocyte differentiation for screening  
XX adipocyte-related disorders such as diabetes and hyperlipemia  
XX  
PS Claim 1; Page 68-72; 112pp; Japanese.  
XX  
CC The invention relates to genes derived from murine adipocytes and  
CC proteins encoded by these genes. The proteins are associated with  
CC adipocyte differentiation and can be expressed by standard recombinant  
CC methodology. The genes, proteins and specific antibodies are useful for  
CC the identification of drugs for treatment and prevention of adipocyte-  
CC related disorders such as obesity, hyperlipemia, diabetes and  
CC atherosclerosis. The present sequence represents a DNA encoding a  
CC protein derived from murine adipocytes.  
XX  
SQ Sequence 1927 BP; 509 A; 536 C; 517 G; 365 T; 0 other;  
  
Query Match 0.8%; Score 20; DB 22; Length 1927;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2276 cctcgggtgccccggagcgc 2295  
|||||  
Db 103 cctcgggtgccccggagcgc 122  
  
RESULT 11  
AAC84391  
ID AAC84391 standard; cDNA; 1949 BP.  
XX  
AC AAC84391;  
XX  
DT 02-APR-2001 (first entry)  
XX

DE Mouse A236 polypeptide encoding cDNA.  
XX  
KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;  
KW antiasthmatic; neuroprotective, cytostatic; cardiant; hepatotropic;  
KW antiinflammatory; antidiabetic; antifertility; antipyretic; vasotropic;  
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;  
KW ophthalmological; antischilling; antiulcer; vulnary; ss.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 304..1425  
FT /\*tag= a  
FT /product= "mouse A236"  
XX  
PN WO200069895-A2.  
XX  
PD 23-NOV-2000.  
XX  
PF 15-MAY-2000; 2000WO-US13361.  
XX  
PR 14-MAY-1999; 99US-0312359.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Pan Y, Leiby KR;  
XX  
DR WPI; 2001-024999/03.  
DR P-PSDB; AAB48126.  
XX  
XX Novel nucleic acids encoding secreted or transmembrane proteins, useful  
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
XX the lung, liver, kidney or pancreas -  
XX  
PS Claim 2; Fig 24A-D; 209pp; English.  
XX  
CC The invention provides human and mouse nucleic acids designated TANGO  
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
CC proteins. The polypeptides, nucleic acids and their modulators may be  
CC useful for treating or modulating cholesterol uptake, blood coagulation,  
CC to modulate cell proliferation, morphogenesis and fate specification,  
CC tissue repair and renewal to treat cancer and promote wound healing,  
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
CC syndrome, protein S deficiency, modulate allergic or inflammatory  
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and  
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
CC herniations, meningitis, ischemic brain or heart disease, infarction,  
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
CC disease, pulmonary heart disease, atherosclerosis, hypertension, jaundice,  
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune  
CC Albright syndrome, infertility, uterine disorders, viral disease. The  
CC present sequence represents the mouse A236 cDNA.  
XX  
SQ Sequence 1949 BP; 508 A; 550 C; 527 G; 364 T; 0 other;

Query Match 0.8%; Score 20; DB 22; Length 1949;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2276 cctcgggtgccccggagcgc 2295  
|||||  
Db 118 cctcgggtgccccggagcgc 137  
  
RESULT 12

```

AAC84416
ID AAC84416 standard; cDNA; 1949 BP.
XX
AC AAC84416;
XX
DT 02-APR-2001 (first entry)
XX
DE Mouse A236 variant 1 cDNA.
XX
KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
KW antiasthmatic; neuroprotective; cytoskeletal; cardiast; hepatotropic;
KW antiinflammatory; antidiabetic; antinfertility; antipyrretic; vasotropic;
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
KW ophthalmological; antiskilling; antiulcer; vulnerary; variant; ss.
XX
OS Mus sp.
XX
FH Key
FH CDS 304..1425
FT /*tag= a
FT /*product= "mouse A236 variant 1"
FT variation 366
FT /*tag= b
FT /*note= "wild-type G at this position is replaced with C"
XX
PN WO200069885-A2.
XX
PD 23-NOV-2000.
XX
PF 15-MAY-2000; 2000WO-US13361.
XX
PR 14-MAY-1999; 99US-0312359.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Pan Y, Leiby KR;
XX
DR WPI; 2001-024999/03.
XX
P-PSDB; AAB48148.
XX
Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
PT the lung, liver, kidney or pancreas -
XX
PS Claim 2; Page -: 209pp; English.
XX
The invention provides human and mouse nucleic acids designated TANGO
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC proteins. The polypeptides, nucleic acids and their modulators may be
CC useful for treating or modulating cholesterol uptake, blood coagulation,
CC to modulate cell proliferation, morphogenesis and fate specification,
CC tissue repair and renewal, to treat cancer and promote wound healing,
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC syndrome, protein S deficiency, modulate allergic or inflammatory
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC herniations, meningitis, ischemic brain or heart disease, infarction,
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC disease, pulmonary heart disease, rheumatic fever, congenital heart
CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents a mouse A236 variant cDNA.
CC Note: the present variant sequence has been constructed using the
CC information provided in the specification.
XX
SQ Sequence 1949 BP; 508 A; 551 C; 526 G; 364 T; 0 other;

```

Query Match 0.88; Score 20; DB 22; Length 1949;  
 Best Local Similarity 100.0%; Pred. NO. 51;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2276 cctcggtgccccggagcgc 2295  
 |||||  
 Db 118 cctcggtgccccggagcgc 137

RESULT 13  
 AAC84417  
 ID AAC84417 standard; cDNA; 1949 BP.

XX AAC84417;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Mouse A236 variant 2 cDNA.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;  
 KW antiasthmatic; neuroprotective; cytoskeletal; cardiast; hepatotropic;  
 KW antiinflammatory; antidiabetic; antinfertility; antipyrretic; vasotropic;  
 KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;  
 KW ophthalmological; antiskilling; antiulcer; vulnerary; variant; ss.

XX Mus sp.  
 XX  
 FH Key  
 FH CDS 304..1425  
 FT /\*tag= a  
 FT /\*product= "mouse A236 variant 2"  
 FT variation 384  
 FT /\*tag= b  
 FT /\*note= "wild-type G at this position is replaced with C"

WO200069885-A2.

23-NOV-2000.

15-MAY-2000; 2000WO-US13361.

14-MAY-1999; 99US-0312359.

(MILL-) MILLENNIUM PHARM INC.

Pan Y, Leiby KR;

WPI; 2001-024999/03.

P-PSDB; AAB48149.

Novel nucleic acids encoding secreted or transmembrane proteins, useful  
 for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
 the lung, liver, kidney or pancreas -

Claim 2; Page -: 209pp; English.

The invention provides human and mouse nucleic acids designated TANGO  
 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
 proteins. The polypeptides, nucleic acids and their modulators may be  
 useful for treating or modulating cholesterol uptake, blood coagulation,  
 to modulate cell proliferation, morphogenesis and fate specification,  
 tissue repair and renewal, to treat cancer and promote wound healing,  
 modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
 syndrome, protein S deficiency, modulate allergic or inflammatory  
 response, acid secretion, tropic effects on gastrointestinal mucosa, and  
 promote ulcer healing, treat bone cancer, achondroplasia, myeloma, and  
 fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
 leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
 herniations, meningitis, ischemic brain or heart disease, infarction,  
 intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart



CC hepatitis C virus (HCV) proteins. The method can be used to identify  
CC compounds which can be used for the prevention and treatment of liver  
CC disorders associated with HCV infection, such as liver cancer. The  
CC present sequence was used in the present invention.

XX  
SQ Sequence 1995 BP; 434 A; 594 C; 635 G; 332 T; 0 other;

Query Match 0.8%; Score 20; DB 22; Length 1995;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2229 ccagcacccgcagctgcagt 2248  
|||||  
Db 1357 CCAGCACCGCGCAGCTGCAGT 1338

Search completed: July 23, 2002, 20:02:27  
Job time: 5707 sec





PT disorders and neurological diseases -

PS Claim 11; Page 274-276; 287pp; English.

XX The present invention relates to seven members of the ADAMS (proteins  
CC which contain A Disintegrin And Metalloprotease domain) protein family.  
CC The ADAMS proteins and DNA may be used to treat disease, as a food  
CC additive or preservative, for chromosome identification, as probes  
CC for diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the  
CC gene therapy of breast and ovarian cancer.

XX Sequence 491 AA;

Query Match 31.7%; Score 1554; DB 22; Length 491;  
Best Local Similarity 99.3%; Pred. No. 6.6e-80;  
Matches 272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 584 MIFQENGVYQYVVISPPPILENPTPEPPVQLOPILVEPLAPAPARTPTGLQ 643  
Db 1 mifqenpvyqyvispppielenptpeppvqlpeillveplapapartptgltq 60  
QY 644 ROVRPQMPAPPHPTPLGSPAAYWKRGHSACSGKGYWRIFLCSRESGEELDER 703  
Db 61 rqvripmpappphprtplsapaaywkrghsaacsckgwrpiflcsresgeelider 120  
QY 704 SCAGARPAPPECHGTPCPYEWAGTWTSCRSCTGTHRQLQCRERGGGSSVPP 763  
Db 121 scaagarpaspechgtcpypweagwtscrsctgthrqqlcqrqefgggssvpp 180  
QY 764 ERGCHLPNITQSCQLRCLGHWEGVSWSCVRCGRQSRQVRCVGNNGDEVSEQC 823  
Db 181 ercghlpnltqscqlrlcghwegvswscvrcgrqsrqvcvngngdevseqc 240  
QY 824 ASGPQPPSREACDMGPTTAWFHDWSKVSPE 857  
Db 241 asgpppsreacdmgpttawfhdwskscae 274

RESULT 2

AAE09713  
ID AAE09713 standard; Protein; 260 AA.

XX AC AAE09713;

XX DT 22-NOV-2001 (first entry)

XX DE Human gene 9 encoding novel protein HUCMO06, SEQ ID NO:60.

XX KW Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
KW autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;  
KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;  
KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
KW anti-sense therapy; endocrine disorder; leukaemia.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

FT Misc-difference 235

FT /label= Unknown

FT /note= "Encoded by TGN"

XX PN WO200155202-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01325.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231442.  
PR 08-SEP-2000; 2000US-0231443.  
PR 08-SEP-2000; 2000US-0231444.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231568.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.

PR	02-OCT-2000;	2000US-0237039.	PS	Claim 11; SEQ ID NO 60; 469pp; English.
PR	13-OCT-2000;	2000US-0237040.	XX	AA016750-AA016775 represent cDNAs corresponding to novel human protein
PR	13-OCT-2000;	2000US-0239935.	CC	genes, and AA009690-AA009715 represent the proteins they encode. AA016777
PR	13-OCT-2000;	2000US-0239937.	CC	AA016780 represent novel human genomic DNA fragments. The novel proteins
PR	20-OCT-2000;	2000US-0240960.	CC	and their DNAs are useful for diagnosing, treating, preventing and/or
PR	20-OCT-2000;	2000US-0241221.	CC	prognosing inflammatory disorders (bursitis or tendonitis); neural
PR	20-OCT-2000;	2000US-0241785.	CC	disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system
PR	20-OCT-2000;	2000US-0241786.	CC	disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);
PR	20-OCT-2000;	2000US-0241808.	CC	muscular disorders; reproductive disorders; gastrointestinal disorders
PR	20-OCT-2000;	2000US-0241809.	CC	(malabsorption syndrome, Crohn's disease); pulmonary disorders;
PR	20-OCT-2000;	2000US-0241826.	CC	cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);
PR	01-NOV-2000;	2000US-0241617.	CC	renal disorders (glomerulonephritis, nephrotic syndrome); cancerous
PR	08-NOV-2000;	2000US-0244774.	CC	disease and conditions (breast cancer); hyperproliferative disorders
PR	08-NOV-2000;	2000US-0244775.	CC	(leukaemia, hyperplasia); tumours; foetal and developmental
PR	08-NOV-2000;	2000US-0246475.	CC	abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,
PR	08-NOV-2000;	2000US-0246476.	CC	asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine
PR	08-NOV-2000;	2000US-0246477.	CC	disorders; pregnancy-related disorders and infections. The novel protein
PR	08-NOV-2000;	2000US-0246478.	CC	DNA is useful in gene therapy and anti-sense therapy. The proteins can
PR	08-NOV-2000;	2000US-0246523.	CC	also be used to aid wound healing and epithelial cell proliferation,
PR	08-NOV-2000;	2000US-0248524.	CC	to prevent skin aging due to sunburn, to maintain organs before
PR	08-NOV-2000;	2000US-0248525.	CC	transplantation, for supporting cell culture of primary tissues, to
PR	08-NOV-2000;	2000US-0248526.	CC	regenerate tissues, to identify their cognate ligands or binding
PR	08-NOV-2000;	2000US-0248527.	CC	
PR	08-NOV-2000;	2000US-0246528.	CC	
PR	08-NOV-2000;	2000US-0246532.	CC	
PR	08-NOV-2000;	2000US-0246609.	CC	
PR	08-NOV-2000;	2000US-0246610.	CC	
PR	08-NOV-2000;	2000US-0246611.	CC	
PR	08-NOV-2000;	2000US-0246613.	CC	
PR	17-NOV-2000;	2000US-0249207.	CC	
PR	17-NOV-2000;	2000US-0249208.	CC	
PR	17-NOV-2000;	2000US-0249209.	CC	
PR	17-NOV-2000;	2000US-0249210.	CC	
PR	17-NOV-2000;	2000US-0249211.	CC	
PR	17-NOV-2000;	2000US-0249212.	CC	
PR	17-NOV-2000;	2000US-0249213.	CC	
PR	17-NOV-2000;	2000US-0249214.	CC	
PR	17-NOV-2000;	2000US-0249215.	CC	
PR	17-NOV-2000;	2000US-0249216.	CC	
PR	17-NOV-2000;	2000US-0249217.	CC	
PR	17-NOV-2000;	2000US-0249218.	CC	
PR	17-NOV-2000;	2000US-0249219.	CC	
PR	17-NOV-2000;	2000US-0249244.	CC	
PR	17-NOV-2000;	2000US-0249245.	CC	
PR	17-NOV-2000;	2000US-0249264.	CC	
PR	17-NOV-2000;	2000US-0249265.	CC	
PR	17-NOV-2000;	2000US-0249297.	CC	
PR	17-NOV-2000;	2000US-0249299.	CC	
PR	17-NOV-2000;	2000US-0249300.	CC	
PR	01-DEC-2000;	2000US-0250160.	CC	
PR	05-DEC-2000;	2000US-0250391.	CC	
PR	05-DEC-2000;	2000US-0251030.	CC	
PR	05-DEC-2000;	2000US-0251988.	CC	
PR	05-DEC-2000;	2000US-0256719.	CC	
PR	06-DEC-2000;	2000US-0251479.	CC	
PR	08-DEC-2000;	2000US-0251856.	CC	
PR	08-DEC-2000;	2000US-0251868.	CC	
PR	08-DEC-2000;	2000US-0251869.	CC	
PR	08-DEC-2000;	2000US-0251989.	CC	
PR	08-DEC-2000;	2000US-0251990.	CC	
PR	11-DEC-2000;	2000US-0254097.	CC	
PR	05-JAN-2001;	2001US-0259678.	CC	
PR	(HUMA-) HUMAN GENOME SCI INC.		XX	
XX	Rosen CA, Barash SC, Ruben SM;		XX	
XX	WPI; 2001-451925/48.		XX	
XX	N-PSDB; AAD16773.		XX	
XX	Isolated polypeptide for treating, preventing and/ or prognosing		XX	
PT	medical disorders and also for testing and detection e.g. diagnosis and		PT	
PT	Screening for agonists -		PT	
XX			XX	

QY	708	GARPPASPEPCHTGTCPPYWEAGWTSCRGCGTQHQROLCRQFEGGSSVPPERC	767
DB	6	gcqapaspechgtbpcpyweagwtscrgcgptqhrqlqcrqfegggssvppercg	65
QY	768	HLPRNITQSCQLRLCGHWEVGPWSQSVRCGRGRQRVRCVGNNGDVEYSEQECASGP	827
DB	66	hlprnitqscqlrlcghwevgsqsvrcgrgrqrgrvrcvgnngdevsequecasgp	125
QY	828	POPPSREACDMGPTATFHSQVSKVSPE	857
DB	126	pqppsreacdmgptatfhsdwskscae	155
RESULT	3		
AA009698			
ID	AA009698	standard; Protein; 372 AA.	
XX	AA009698;		
XX	22-NOV-2001	(first entry)	
DE	Human gene 9	encoding novel protein HUCMO06, SEQ ID NO:45.	
XX	Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;		
XX	Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;		
XX	autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;		
XX	reproductive disorder; Crohn's disease; pulmonary disorder; cancer;		
XX	myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;		
XX	haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;		
XX	anti-sense therapy; endocrine disorder; leukaemia.		
XX	Homo sapiens.		
OS	WO200155202-A1.		
PN	02-AUG-2001.		
XX	17-JAN-2001; 2001WO-US01325.		
XX	31-JAN-2000; 2000US-0179065.		
XX	04-FEB-2000; 2000US-0180628.		
XX	24-FEB-2000; 2000US-0184664.		
XX	02-MAR-2000; 2000US-0186350.		
XX	16-MAR-2000; 2000US-0189874.		
XX	17-MAR-2000; 2000US-0190076.		

PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0218293.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0227109.  
PR 23-AUG-2000; 2000US-0227109.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239535.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249284.  
PR 17-NOV-2000; 2000US-0249285.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0255678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-451925/48.  
XX N-PSDB; AAD16758.  
XX Isolated polypeptide for treating, preventing and/or prognosing  
XX medical disorders and also for testing and detection e.g. diagnosis and  
XX screening for agonists -  
XX Claim 11; SEQ ID No 45; 469pp; English.  
XX AAD16750-AAD16775 represent cDNAs corresponding to novel human protein  
XX genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777  
CC

CC -AAD16780 represent novel human genomic DNA fragments. The novel proteins  
CC and their DNAs are useful for diagnosing, treating, preventing and/or  
CC prognosing inflammatory disorders (bursitis or tendonitis); neural  
CC disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system  
CC disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);  
CC muscular disorders; reproductive disorders; gastrointestinal disorders  
CC (malabsorption syndrome, Crohn's disease); pulmonary disorders;  
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);  
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous  
CC disease and conditions (breast cancer); hyperproliferative disorders  
CC (leukaemia, hyperplasia); tumours; foetal and developmental  
CC abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,  
CC asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine  
CC disorders; pregnancy-related disorders and infections. The proteins can  
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can  
CC also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemotaxis, and can be used as a food additive or  
CC preservative to modify storage properties. The present sequence  
CC represents a novel human protein of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly

Query Match 17.5%; Score 858; DB 22; Length 372;  
Best Local Similarity 96.7%; Pred. No. 3.7e-45;  
Matches 145; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 708 GARPASPEPCGHPCCPYWAGETSCSRSCGPGTQHQRLQCRQEGCGSSVPPERC 767  
DB 6 GCPAPSPCHGTPCPCPYWAGETSCSRSCGPGTQHQRLQCRQEGCGSSVPPERC 65  
QY 768 HLPNRTQSCQLRCLGHWEGSPWQSVRCGRQSRQVRCVGNNGDEVSEQECASGP 827  
DB 66 HLPNRTQSCQLRCLGHWEGSPWQSVRCGRQSRQVRCVGNNGDEVSEQECASGP 125  
QY 828 POPPSREACDMGPCTATFPHSDWSKVSPE 857  
DB 126 POPPSREACDMGPCTATFPHSDWSKVSPE 155

RESULT 4  
AAB50947  
ID AAB50947 standard; peptide; 150 AA.  
XX  
AC AAB50947;  
XX  
DT 20-MAR-2001 (first entry)  
XX  
DE ADAM gene #2 peptide.  
XX  
KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;  
KW ovarian.  
XX  
OS Homo sapiens.  
XX  
FN WO200073323-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 25-MAY-2000; 2000WO-US14308.  
XX  
PR 27-MAY-1999; 99US-0136388.  
PR 09-JUL-1999; 99US-0142930.  
PR 28-JAN-2000; 2000US-0176717.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;  
XX  
DR WPI; 2001-016507/02.

PT Seven nucleic acid molecules encoding ADAM polypeptides containing a  
PT disintegrin and metalloprotease domain, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
XX disorders and neurological diseases.  
XX  
XX Disclosure; Page 17; 287pp; English.  
XX  
XX The present invention relates to seven members of the ADAMs (proteins  
CC which contain A Disintegrin And Metalloprotease domain) protein family.  
CC The ADAMs proteins and DNA may be used to treat disease, as a food  
CC additive or preservative, for chromosome identification, as probes  
CC for diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovarian cancer. They are also useful in the  
CC gene therapy of breast and ovarian cancer.  
XX  
XX Sequence 150 AA;

Query Match 16.7%; Score 817; DB 22; Length 150;  
Best Local Similarity 100.0%; Pred. No. 4.5e-43;  
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 435 EKQDGTLCOPGAPDICVAGRCCLSPGCDGILGSGRRPDCGCGGDDSTCRIVSGNLTR 494  
DB 2 EKQDGTLCOPGAPDICVAGRCCLSPGCDGILGSGRRPDCGCGGDDSTCRIVSGNLTR 61  
QY 495 GGPLGYQKILWIPAGALRLQIALRPSSNYLALRPGGRRSIINGNWAYDPPGSRAGGTV 554  
DB 62 GGPLGYQKILWIPAGALRLQIALRPSSNYLALRPGGRRSIINGNWAYDPPGSRAGGTV 121  
QY 555 FRYNRPREGKESLSAEGTTPQVDVY 583  
DB 122 FRYNRPREGKESLSAEGTTPQVDVY 150

RESULT 5  
AAB6949  
ID AAB6949 standard; Protein; 1690 AA.  
XX  
AC AAB6949;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human metalloprotease MPTS-19 protein.  
XX  
KW Metalloprotease; MPTS-19; human; thrombospondin domain; spondylitis;  
KW aggrecan associated disease rheumatoid arthritis; osteoarthritis;  
KW osteopathic; antiarthritic; antipsoriatic; antirheumatic; psoriasis;  
KW sport injury; joint trauma; fibrosis.  
XX  
OS Homo sapiens.  
XX  
FN DE10107360-A1.  
XX  
PD 06-SEP-2001.  
XX  
PF 16-FEB-2001; 2001DE-1007360.  
XX  
PR 18-FEB-2000; 2000US-184152P.  
XX  
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
PI Heller RA, Klonowski P, Zuo F;  
XX  
DR WPI; 2001-607955/70.  
DR N-PSDB; AAH49372.  
XX  
PT Novel metalloprotease containing a thrombospondin domain (MPTS protein)  
PT is useful to treat aggrecan associated disease including rheumatoid  
PT arthritis and osteoarthritis.  
XX  
PS Claim 1a; Page 33-38; 66pp; German.  
XX

CC This invention describes a novel metalloprotease containing a  
 CC thrombospondin domain (MPS protein) which is useful to treat aggrean  
 CC associated disease including rheumatoid arthritis and osteoarthritis.  
 CC The products of the invention have osteopathic, antiarthritic,  
 CC antiproliferative and antirheumatic activity. MPS may be used to identify a  
 CC modulator of its activity, e.g., an agonist or antagonists. Such  
 CC compounds, as well as MPS itself may be used to treat disease  
 CC associated with MPS activity or aggrean degradation, particularly  
 CC osteoarthritis, rheumatoid arthritis, psoriasis, spondylitis, sport  
 CC injury, joint trauma or fibrosis. This sequence represents the human  
 CC MPS-19 metalloprotease described in the method of the invention.  
 XX  
 SQ Sequence 1690 AA;

Query Match 15.18; Score 740.5; DB 22; Length 1690;  
 Best Local Similarity 26.08; Pred. No. 3.4e-37;  
 Matches 245; Conservative 94; Mismatches 298; Indels 305; Gaps 39;

QY 86 PRPPRH-----PEALLPRGQGRPOTS-----PETPLRYTQSRGGGRLRGA 129  
 DB 176 parpghaiphvyrkqaperlaqgdasapstcgvyypele----- 218  
 QY 130 SHLGRETEITRAARRSLRPIKPMFGYGRVFPALPLHRRNRHRSPPR----- 180  
 DB 219 ---rrerweqrgwrplr-----rllhgrsvskewwetlrv 253  
 QY 181 SELSLISSRGEAIPSPRAEPF-----SANGSP-----QT 212  
 DB 254 adaknveyhgq-----pqvesyvtlmmnavglfhdpisgnhlitvrlvledee 306  
 QY 213 ELPPTELSVHTPSP-----QAEPLSPETA-----QTEVAPRTRPAP---LRH- 251  
 DB 307 dlkithadnptkscfkqsknmkgdahphhdtailtrkdkatnmrncetlgishv 366  
 QY 252 ----HPRQAQSTEP-----SPHSLG-----EGGFRASPQPRPSSQGNASPOV 294  
 DB 367 agmcqphrscsinedtqlaftvaheghsfqhdgsgndcepvgkrpfim---spql 423  
 QY 295 -----AGRR-----PDP-----FPSVRG-----RGOQG 313  
 DB 424 lydaapltwscrgyitrflrgrwglclddpakdlidfpvpgvlydvshqrlqvg 483  
 QY 314 -----QGPWGTGTPHPRLPPQH-----PGAWLPL-----LSN 344  
 DB 484 aysafcedmndvchtlwcsvgttchskidaavdgtrognkwcslgcevpvgrfpeavdg 543  
 QY 345 GHASLSLWLFAPS-----SPIPR-----CSGESEQLRACSOAPCPPEQDP 386  
 DB 544 gswgswawsicrscgmvgqaeqrctqtpkygrycvgerkrfrlcnlqacpagrpsf 603  
 QY 387 RALQCAAFNSQEFMQLQWEPFTEVQGSQRCNCRPRGFRFYVHTEKVDGTLQCP- 445  
 DB 604 rhvqcshtdamllykgrlhtwvrv---vndvnpcehlcrpaneyfaeklrдавvdgtpcyqv 661  
 QY 446 -GAPDICVAGRLSPGCGILGSGRRPDCGCGGDDSTCLVSLNLTDRGGLGYKXIL 504  
 DB 662 rasrdclngicknvcdfeldsgamedrcgchngstchtvsgtfeaaq-lgyvdvg 720  
 QY 505 WTPAGALRLQIAQLRPSNYLALRFG-GRSLINGNNAVDPGSRAGTGYFRNRPPE 563  
 DB 721 lllpagaerlqdeavaeanflalrsdebekyflngdqlqwnvgdvgvgaqftfyar----- 776  
 QY 564 EKGGSLSAEGPTTQPDVMTFQSENPCVFYQYVSSPPPLENPTPEPPVPVLPQPEIL 623  
 DB 777 rgnwvnltspgtkpevniqlifqesngpvhyey----- 810  
 QY 624 RVEPFLAPRPARTEGTLQORVIRIQPAPPHPTPLGSPAAYWKRVGHSACSASCKG 683  
 DB 811 -----tltre-----agghdevp---ppvfwshyggwtckctcgrg 844  
 QY 684 VWRPIFLCTSRSGBELDERSAAGARRPASPEPCHGTPCPYPWEAGWTCSRSCGP-G 742

DB 845 vqrqnyclierqag-pvdeehcdplgrpdqqrkscqpcparwagewqlcssscgp99 903  
 QY 743 TOHRQLQCRQFEG-GGGSVPPERCCHLRPNITQSCQLRL-C-GHWEVGSFWSQCSVRC 799  
 DB 904 lsrravlcirsvgldeqsaleppacehlprrptetpcnrhvpccpatwvgn-wsqcsvrc 962  
 QY 800 GRGQRSRQVRVCVGNNDDEVSEOCASGPPPPSREACDMGPC 841  
 DB 963 gegtqrnvlctndtgyvpcdeag-----qpasevtctslpc 998  
 RESULT 6  
 AAE00934  
 ID AAE00934 standard; Protein; 1686 AA.  
 AC AAE00934;  
 XX  
 DT 04-JUL-2001 (first entry)  
 XX  
 DE Human 27875 ADAM-TS (a disintegrin and metalloproteinase).  
 KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antinflammatory;  
 KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;  
 KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;  
 KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;  
 KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;  
 KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;  
 KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;  
 KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;  
 KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;  
 KW cardiant; tumour; thymoma; vasotropic; cytostatic; virucide.  
 XX Homo sapiens.  
 OS  
 PH  
 FT Key Location/Qualifiers  
 FT Peptide 1..30  
 FT /label= Signal\_peptide  
 FT Modified-site 6..8  
 FT /note= "Protein kinase C phosphorylation site"  
 FT Protein 31..1686  
 FT /label= Human\_mature\_27875\_ADAM-TS\_protein  
 FT Modified-site 55..60  
 FT /note= "N-myristoylation site"  
 FT Modified-site 73..75  
 FT /note= "Protein kinase C phosphorylation site"  
 FT Domain 78..93  
 FT /note= "Crystallins beta and gamma Greek key motif"  
 FT Modified-site 94..97  
 FT /note= "N-glycosylation site"  
 FT Modified-site 110..112  
 FT /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 115..120  
 FT /note= "N-myristoylation site"  
 FT Modified-site 141..146  
 FT /note= "N-myristoylation site"  
 FT Modified-site 147..150  
 FT /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 159..162  
 FT /note= "Casein kinase II phosphorylation site"  
 FT Binding-site 195..197  
 FT /note= "Cell attachment sequence"  
 FT Modified-site 214..217  
 FT /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 214..216  
 FT /note= "Protein kinase C phosphorylation site"  
 FT Domain 244..259  
 FT /label= Metalloproteinase\_domain  
 FT Modified-site 313..315  
 FT /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 342..345  
 FT /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 342..344  
 FT /note= "Protein kinase C phosphorylation site"

```

FT Modified-site 373..376 /note="Casein kinase II phosphorylation site"
FT Modified-site 379..384 /note="N-myristoylation site"
FT Domain 385..394 /label= Zinc_binding_domain
FT Modified-site 401..404 /note="Casein kinase II phosphorylation site"
FT Modified-site 408..411 /note="Amidation site"
FT Modified-site 479..484 /note="N-myristoylation site"
FT Domain 488..567 /label= Thrombospondin_domain
FT Modified-site 505..508 /note="Casein kinase II phosphorylation site"
FT Modified-site 513..518 /note="N-myristoylation site"
FT Domain 539..545 /note="Growth factor and cytokine metalloproteinase family signature 2 domain"
FT Modified-site 539..544 /note="N-myristoylation site"
FT Domain 541..592 /label= Disintegrin_domain
FT Domain 542..592 /label= Thrombospondin_domain
FT Modified-site 557..562 /note="N-myristoylation site"
FT Modified-site 569..571 /note="Protein kinase C phosphorylation site"
FT Modified-site 598..600 /note="Protein kinase C phosphorylation site"
FT Modified-site 605..608 /note="Casein kinase II phosphorylation site"
FT Modified-site 614..619 /note="N-myristoylation site"
FT Modified-site 667..672 /note="N-myristoylation site"
FT Binding-site 687..692 /note="Cytochrome C family heme-binding site"
FT Modified-site 688..693 /note="N-myristoylation site"
FT Modified-site 693..696 /note="N-glycosylation site"
FT Modified-site 703..706 /note="Casein kinase II phosphorylation site"
FT Modified-site 716..721 /note="N-myristoylation site"
FT Modified-site 740..747 /note="Tyrosine kinase phosphorylation site"
FT Modified-site 765..770 /note="N-myristoylation site"
FT Modified-site 774..779 /note="N-myristoylation site"
FT Modified-site 778..781 /note="N-glycosylation site"
FT Domain 825..868 /label= Thrombospondin_domain
FT Modified-site 872..875 /note="cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 901..903 /note="Protein kinase C phosphorylation site"
FT Modified-site 917..920 /note="Casein kinase II phosphorylation site"
FT Domain 949..994 /label= Thrombospondin_domain
FT Domain 949..988 /label= Thrombospondin_domain
FT Modified-site 950..953 /note="N-glycosylation site"
FT Modified-site 957..960

```

```

FT Modified-site 962..964 /note="Casein kinase II phosphorylation site"
FT Modified-site 971..974 /note="Protein kinase C phosphorylation site"
FT Modified-site 1005..1010 /note="N-glycosylation site"
FT Binding-site 1006..1009 /note="N-myristoylation site"
FT Modified-site 1011..1014 /note="Glycosaminoglycan attachment site"
FT Modified-site 1035..1037 /note="Casein kinase II phosphorylation site"
FT Modified-site 1039..1044 /note="Protein kinase C phosphorylation site"
FT Modified-site 1192..1195 /note="N-myristoylation site"
FT Modified-site 1263..1252 /note="Casein kinase II phosphorylation site"
FT Modified-site 1263..1268 /note="N-myristoylation site"
FT Modified-site 1308..1311 /note="N-myristoylation site"
FT Modified-site 1358..1363 /note="Casein kinase II phosphorylation site"
FT Modified-site 1370..1372 /note="N-myristoylation site"
FT Modified-site 1385..1387 /note="Protein kinase C phosphorylation site"
FT Modified-site 1397..1400 /note="Protein kinase C phosphorylation site"
FT Modified-site 1412..1415 /note="Casein kinase II phosphorylation site"
FT Domain 1415..1463 /note="N-glycosylation site"
FT Modified-site 1419..1422 /label= Thrombospondin_domain
FT Modified-site 1440..1443 /note="N-glycosylation site"
FT Modified-site 1440..1442 /note="Casein kinase II phosphorylation site"
FT Modified-site 1466..1521 /note="Protein kinase C phosphorylation site"
FT Modified-site 1470..1473 /label= Thrombospondin_domain
FT Modified-site 1483..1486 /note="N-glycosylation site"
FT Modified-site 1483..1485 /note="Casein kinase II phosphorylation site"
FT Modified-site 1517..1522 /note="Protein kinase C phosphorylation site"
FT Modified-site 1528..1531 /note="N-myristoylation site"
FT Modified-site 1528..1530 /note="Casein kinase II phosphorylation site"
FT Modified-site 1546..1549 /note="Protein kinase C phosphorylation site"
FT Modified-site 1592..1597 /note="Casein kinase II phosphorylation site"
FT Modified-site 1599..1601 /note="N-myristoylation site"
FT Modified-site 1606..1609 /note="Protein kinase C phosphorylation site"
FT Modified-site 1606..1609 /note="cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 1620..1622 /note="Protein kinase C phosphorylation site"
FT Modified-site 1625..1630 /note="Protein kinase C phosphorylation site"

```

Query Match 15.0%; Score 736.5; DB 22; Length 1686;  
 Best Local Similarity 26.0%; Pred. No. 6e-37;

Matches 245; Conservative 95; Mismatches 297; Indels 305; Gaps 40;

QY 86 PRPRH-----PEALLPRGQGRPQTS-----PETLPLYQTQSRGGPLRGPA 129



```
Db 172 parphagphvvykrqaperlaqrdsasbstcgqvvyype-----lesr----- 215
QY 130 SHLGRETEQIEIARAARSLRDPKFGMEGYGRVPFALPLHRNRHPRPPR----- 180
Db 216 -----terweqgqrwrplr-----rlhqrsvskekwetlrv 249
QY 181 SELSLISSRGEAIPSPTRAPPP-----SANGSP-----QTELP 215
Db 250 adakmveyhgq-----pqvesyvltnmmvagi fhdpsignphltivrlvledee 302
QY 216 PTELSVHTPSP-----QAEPLSPETA-----QTEVAPTRPAP---LRH- 251
Db 303 dlkithadntlksckwqksimkgdahplhdtailtrkdcaamnrpoetlglshv 362
QY 252 ----HPRQAQSGTEPP-----SPTHSLG-----EGGFFRASPPRRPSQGWASQV 294
Db 363 agmcqphrscsinedtglplafvahalghsfghdgsndcepvgkrpfim---spql 419
QY 295 -----AGRR-----PDP-----PPSVPRG-----RGQOG 313
Db 420 lydaaplwtwscrsqyitrlfdrgwglclddpakdiidfpvpgvlydvshqcrlyg 479
QY 314 -----QGPWGTGTPHGRPLEDPQH-----PGAWLPL-----LSN 344
Db 480 aysafcedmndvchtlwcsvgttchskldaavdgtrcgcnkwcisgcevpvgrpeavdg 539
QY 345 GPHASSLSLWLFAPS-----SPIPR-----CSGESEQLRACSOACPPPPQDP 386
Db 540 gwsawsicrsrscgmvgvaerqctqtpckykrycvgerkrfrlcnlqacpagrpsf 599
QY 387 RALQCAAFNSQEFMQLYQWEPFTEVQGSQRCCELNCRPRGFYVRHTEKVDQGTLCQP- 445
Db 600 rhvqcsfhdamlykgqlhtwvvp--vndvnpcelhcrpaneyfaeklrдавдgtpcyqv 657
QY 446 -CAPDICVAGRLSPGCGILGSGRRPGCGVCGGDDSTCLVSNLNDRGGLYQKIL 504
Db 658 rsrdclngicnkvocfidsamedrcgvchngstchtvstfgeaag-lyyvdvg 716
QY 505 WTPAGALRLQIAQLRPSSNYIALRPG--GRSLINGNNAVDPPGSYRAGTVFRYNNRPPRE 563
Db 717 lllpagareiridevaeaanflalrsedbekyflngwtiqwngdyqvagttftfvar----- 772
QY 564 EKGGSLSAEGTTPQVDVYIMIFQENPVFQYVVISPPPLENPTPEPPVLPQLPBIL 623
Db 773 rgnwlnltpgptkpevwiqlifqesnpgvhyey----- 806
QY 624 RVEPPLAPAPRPTPTGLQKQVRIQPMAPPHPTPLGLSPPAAYKVRVGHGSACSCKG 683
Db 807 -----tlhre-----agghdevp--ppvfwshygpwtkctvtcgr 840
QY 684 VWRPFIICISRESGELDERSCAAGRPASPPECHGTPCYPYNEAGETWCSRSCKG-G 742
Db 841 vqrnvycrlcrqag-pvdeehcdplgrddqgrkxscqpcparwagewqlcscscgpgg 899
QY 743 TQHQLOCRQEFQ-GGGSVPPPERGCHLPRPNITQSCQLRL-C-GHWEVGSPPWQCSYRC 799
Db 900 lrravlcirsvgldeqaleapacehlprptetpcnrhvcpatwavn-wsqcstv 958
QY 800 GRGQSRQVRQVGNNGDEVDSQECASGPPQPSRACDMGFC 841
Db 959 gegtgrnvlctndtgvpcdeaq-----qpasevtcslpic 994

RESULT 7
AAB74944
ID AAB74944 standard; Protein; 1686 AA.
XX
AC AAB74944;
XX
XX 02-JUL-2001 (first entry)
XX
DE Human ADAM type metal protease MDTs1 protein SEQ ID NO:1.
```

```
XX
KW Human; a disintegrin and metalloprotease type metal protease; MDTs1;
KW MDTs2; MDTs3; ADAM type metal protease; cytosolic; antiarthritic;
KW cancer; arthritis; arthrosis deformans.
XX
OS Homo sapiens.
XX
PN JP2001008687-A.
XX
PD 16-JAN-2001.
XX
PF 25-JUN-1999; 99JP-0180973.
XX
PR 25-JUN-1999; 99JP-0180973.
XX
PA (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
XX WPI; 2001-285362/30.
XX
XX N-PSDB; AAF82149.
XX
XX New metal protease and metal protease gene, for use as a drug for
XX cancers, arthritis and arthrosis deformans.
XX
XX Claim 1; Page 12-17; 31pp; Japanese.
XX
XX The present sequence represents a disintegrin and metalloprotease (ADAM)
XX type metal protease designated MDTs1, isolated from human. MDTs proteins
XX have cytosolic and antiarthritic activities. They can be used as a drug
XX for cancers, arthritis and arthrosis deformans.
XX
XX Sequence 1686 AA;
```

Query Match 15.0%; Score 735.5; DB 22; Length 1686;  
Best Local Similarity 26.0%; Pred. No. 6.9e-37;  
Matches 245; Conservative 94; Mismatches 298; Indels 305; Gaps 40;

```
QY 86 PRPRH-----PEALLRGGQPRPQTS-----PETLPLYTQSRGGPLRGPA 129
Db 172 parphagphvvykrqaperlaqrdsasbstcgqvvyype-----lesr----- 215
QY 130 SHLGRETEQIEIARAARSLRDPKFGMEGYGRVPFALPLHRNRHPRPPR----- 180
Db 216 -----terweqgqrwrplr-----rlhqrsvskekwetlrv 249
QY 181 SELSLISSRGEAIPSPTRAPPP-----SANGSP-----QTELP 215
Db 250 adakmveyhgq-----pqvesyvltnmmvagi fhdpsignphltivrlvledee 302
QY 216 PTELSVHTPSP-----QAEPLSPETA-----QTEVAPTRPAP---LRH- 251
Db 303 dlkithadntlksckwqksimkgdahplhdtailtrkdcaamnrpoetlglshv 362
QY 252 ----HPRQAQSGTEPP-----SPTHSLG-----EGGFFRASPPRRPSQGWASQV 294
Db 363 agmcqphrscsinedtglplafvahalghsfghdgsndcepvgkrpfim---spql 419
QY 295 -----AGRR-----PDP-----PPSVPRG-----RGQOG 313
Db 420 lydaaplwtwscrsqyitrlfdrgwglclddpakdiidfpvpgvlydvshqcrlyg 479
QY 314 -----QGPWGTGTPHGRPLEDPQH-----PGAWLPL-----LSN 344
Db 480 aysafcedmndvchtlwcsvgttchskldaavdgtrcgcnkwcisgcevpvgrpeavdg 539
QY 345 GPHASSLSLWLFAPS-----SPIPR-----CSGESEQLRACSOACPPPPQDP 386
Db 540 gwsawsicrsrscgmvgvaerqctqtpckykrycvgerkrfrlcnlqacpagrpsf 599
QY 387 RALQCAAFNSQEFMQLYQWEPFTEVQGSQRCCELNCRPRGFYVRHTEKVDQGTLCQP- 445
Db 600 rhvqcsfhdamlykgqlhtwvvp--vndvnpcelhcrpaneyfaeklrдавдgtpcyqv 657
```



FT Modified-site /note= "N-myristoylation site"  
FT 774..779  
FT /note= "N-myristoylation site"  
FT 778..781  
FT /note= "N-glycosylation site"  
FT 825..868  
FT /label= Thrombospondin\_domain  
FT Modified-site  
FT 872..875  
FT /note= "cAMP- and cGMP-dependent protein kinase  
FT phosphorylation site"  
FT 901..903  
FT /note= "Protein kinase C phosphorylation site"  
FT 917..920  
FT /note= "Casein kinase II phosphorylation site"  
FT 949..994  
FT /label= Thrombospondin\_domain  
FT 949..988  
FT /label= Thrombospondin\_domain  
FT Modified-site  
FT 950..953  
FT /note= "N-glycosylation site"  
FT Modified-site  
FT 957..960  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 962..964  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site  
FT 971..974  
FT /note= "N-glycosylation site"  
FT Modified-site  
FT 1005..1010  
FT /note= "N-myristoylation site"  
FT 1006..1009  
FT /note= "Glycosaminoglycan attachment site"  
FT Modified-site  
FT 1011..1014  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 1035..1037  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site  
FT 1039..1044  
FT /note= "N-myristoylation site"  
FT Modified-site  
FT 1192..1195  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 1263..1252  
FT /note= "N-myristoylation site"  
FT Modified-site  
FT 1263..1268  
FT /note= "N-myristoylation site"  
FT Modified-site  
FT 1308..1311  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 1358..1363  
FT /note= "N-myristoylation site"  
FT Modified-site  
FT 1370..1372  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site  
FT 1385..1387  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site  
FT 1397..1400  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 1412..1415  
FT /note= "N-glycosylation site"  
FT 1415..1463  
FT /label= Thrombospondin\_domain  
FT Modified-site  
FT 1419..1422  
FT /note= "N-glycosylation site"  
FT Modified-site  
FT 1440..1443  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 1440..1442  
FT /note= "Protein kinase C phosphorylation site"  
FT 1466..1521  
FT /label= Thrombospondin\_domain  
FT Modified-site  
FT 1470..1473  
FT /note= "N-glycosylation site"  
FT Modified-site  
FT 1483..1486  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site  
FT 1483..1485  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site  
FT 1517..1522  
FT /note= "N-myristoylation site"  
FT 1528..1531

FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT 1528..1530  
FT /note= "Protein kinase C phosphorylation site"  
FT 1546..1549  
FT /note= "Casein kinase II phosphorylation site"  
FT 1592..1597  
FT /note= "N-myristoylation site"  
FT 1599..1601  
FT /note= "Protein kinase C phosphorylation site"  
FT 1606..1609  
FT /note= "cAMP- and cGMP-dependent protein kinase  
FT phosphorylation site"  
FT  
  
Query Watch 15.0%; Score 734.5; DB 22; Length 1686;  
Best Local Similarity 26.0%; Pred. No. 8e-37;  
Matches 245; Conservative 94; Mismatches 298; Indels 305; Gaps 40;  
  
QY 86 PREPRH-----PEALLPRGGQPRPQTS-----PETPLRYTQSRGGPLRQPA 129  
Db 172 parpghaqphvykrqaperlaqrdsaspsstcgqvpe-----lesr----- 215  
QY 130 SHLGREETQIRAAERSRLRDPKPGMGYGRVPPFALPLHRRNRHPRSPPR----- 180  
Db 216 ----rerwedqrqwrprlr-----rlhqrsvskewvetllv 249  
QY 181 SELSLISSRGEAAIPSPTPRAEPF-----SANGSP-----QTLP 215  
Db 250 adakmveyhq-----pqvesyvitimmvaglfhdpsignphtivirvlledee 302  
QY 216 PTLSVHTPSP-----QAEPLSPETA-----QTEVAPRTRPAP-----LRH- 251  
Db 303 dlkithadntlksfckwksimkgdanplhdhtalllrkldcaamrpoetlqshv 362  
QY 252 ----HPRAAQASGTEPP-----SPTHSLG-----EGGFFRASPPRRPSGGWASQV 294  
Db 363 agmcqphrscsinedtqlaftvaheighsfqlqhdgsgndcepvgkrpfm---spql 419  
QY 295 -----AGRR-----PDP-----PPSVPRG-----RGQG 313  
Db 420 lydaapltwrcsrqytrfldrgwglclddpakdiidpsvppgyldvshqclqyg 479  
QY 314 -----QGPWGTGTPHGRPLEDPQH-----PGAWLPL-----LSN 344  
Db 480 aysafcedmrvchtlwcvsgtlychskldaadvgttrcgenkwcslgscvpgfirpeadv 539  
QY 345 GPHASSLWSLPAPS-----SPIPR-----CSGESEQLRACSAOPCPQPDP 386  
Db 540 gwsawsawscscrgmgvqsaerqctqtpkykgrycvgerkrifclnlaqcapgxpsf 599  
QY 387 RALCAAFNSQEFMQLYQWEPFTEVQGSQRCNELNCRPRGFRFYVRHTEKVDGTLQCP- 445  
Db 600 rhvqshfdamlykqqlhtwvpv---vndvnpclhcrpaneyfaeklrдавvgttcyqv 657  
QY 446 -GAPDICVAGRCLSPGCGDILGSGRRRDPDGGVCGGDDSTCLVSGNLTDRGGPLGYKIL 504  
Db 658 rasrdclcingicknvgcdfeldsgamedrcgvchngstchtvsgtfeaeag-lgyvdivg 716  
QY 505 WIPAGALRLQIAOLRPSNNYLALRPG-GRSINGNAVDPPGYSYRAGGTVFYRNRPPRE 563  
Db 717 lllpgareirfqvaeaanfalsedekyflnggwtiqngdyqvagtftyar---- 772  
QY 564 EGKGESLSAEGTTQPDVYVMIFQENPGVYQYVVISPPPILENPEPPVPVLOLPEIL 623  
Db 773 rgnwenltspgtkepvwigilfgesnpgvhye----- 806  
QY 624 RVEPPLAPAPARTPTGLQROVRPQMPAPHPRTPLGSPAAVYKRWKVGHSACSASCGKG 683  
Db 807 -----tjhre-----agghdevp---ppvfshwgygwtckvtcgrg 840  
QY 684 VWRPFLCISRESGEELDERSCAAGAPPASPPECHGTPTCPYWEAGETWTSRSCGPG-G 742  
Db 841 vqrqnvyclerqag-pvdeehcdplgrpddqgrkoseqpcparwwagewqlcscscpgg 899

QY 743 TOHRQLQCRQFG-GGGSSVPPERCGLHLPNPTTQSCQLRL-C-GHWEVGSPPWSQCSVRC 799  
Db 900 lstravlicrsvgldegalepachelprptetpcnhrhvpapwavn-wsgcgvic 958  
QY 800 GRGORSQVRCVNGNDEVSEQCASGPPROPSPREACDMGPC 841  
Db 959 gegtqrnvictndtgvpcdeaq-----qpasevtcsiplc 994  
RESULT 9  
AAU01291  
ID AAU01291 standard; Protein; 634 AA.  
AC AAU01291;  
XX  
XX 18-JUL-2001 (first entry)  
XX Human Thrombospondin repeat domain protein 1, TSR1.  
XX  
XX Human; Thrombospondin repeat domain; TSRX: cancer; breast cancer;  
KW rheumatoid arthritis; ocular neovascularisation; wound healing;  
KW angiogenesis; immune associated disorder; gestational disorder;  
KW pre-eclampsia; neuronal development; immunogen; antibody; antisen-  
KW agonist; TSR1.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..21  
FT /label= Signal\_peptide  
FT 22..634  
FT Protein /label= Mature\_TSR1  
FT Domain 135..185  
FT /label= Complement\_precursor\_repeat\_domain  
FT Domain 140..190  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 418..473  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 477..533  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 535..591  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 596..630  
FT /label= Thrombospondin\_repeat\_domain  
XX  
XX WO200123561-A2.  
XX  
XX  
XX 05-APR-2001.  
XX  
XX 27-SEP-2000; 2000WO-US26432.  
XX  
XX 27-SEP-1999; 99US-0156217.  
XX 27-JUN-2000; 2000US-0214759.  
XX 26-SEP-2000; 2000US-0669360.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shinkets RA, Vernet C, Tchernev VT, Boldog FL, Herrmann JL;  
XX  
XX WPI: 2001-266157/27.  
XX N-PSDB; AA502432.  
XX  
XX TSRX PRO (PRO comprising thrombospondin-1 repeat) domain useful to  
PT identify molecules modulating TSRX activity or function, for treating  
PT cancer, rheumatoid arthritis and ocular neovascularisation -  
XX  
XX Claim 1; Page 6-9; 116pp; English.  
XX  
XX The sequence represents Human thrombospondin-1 repeat (TSR) domain  
CC containing protein, TSR1. Members of the TSR superfamily, TSRX proteins,  
CC include proteins responsible for cell attachment, spreading, motility,  
CC proliferation, cytoskeletal organisation, wound healing and  
CC angiogenesis. TSRX, TSRX polynucleotides and anti-TSRX antibodies are

CC useful for diagnosing, treating or preventing cancer, rheumatoid  
CC arthritis, ocular neovascularisation, wound healing, immune associated  
CC disorders and gestational diseases (e.g. pre-eclampsia). TSRX and TSRX  
CC polynucleotides can be used to identify members of the TSR superfamily,  
CC to screen for molecules which inhibit or enhance TSRX activity or  
CC function, as targets for identification of small molecules that modulate  
CC or inhibit e.g. angiogenesis or neuronal development. Also TSRX antisense  
CC molecules or other agonists are useful for detecting and treating breast  
CC cancer. TSRX proteins can be used to screen drugs or compounds that  
CC modulate TSRX activity or expression as well as to treat disorders  
CC characterised by insufficient or excessive production of TSRX or  
CC production of TSRX forms that have decreased or aberrant activity  
CC compared to TSRX wild-type. Anti-TSRX antibodies can be used to isolate  
CC TSRXs and modulate TSRX activity. Portions or fragments of TSRX cDNAs  
CC are used as polynucleotide reagents and are used for tissue typing  
CC and forensic identification.  
XX  
XX Sequence 634 AA;  
QY 176 RSPRSELSLSRGE-----AIPSTPRAEPFANGSPQTELPPTLSVHTPSQAEP 231  
Db 37 repgrehralcsgtsepdwlp-----ra-----gcpreegrgaslgh-----l 79  
QY 232 SPETAOTEVAPRTRPAPLRHHPRAQASGTPEPTSHSLGEGGFFRASPPRPSQGWAS 291  
Db 80 gpdevselwclsknrcitnsipaagt--lcqthti-----dkgwc 121  
QY 292 PQVAGRRPDPFPVPRG-RGQCGQ-GFWGTGTPHGPRLPEPDQHPGAWLPILLSNGPHAS 349  
Db 122 krvc-----vpfgrpegvdgawgtpwgdcsrtcgvgvsssrhcdsprtigg----- 172  
QY 350 SLMSLPAPSSPIPRCGESQLRACQAPCPPEQPPRALQCAAFNSQFMGLYQWEPF 409  
Db 173 -----xycigerrhrscntddcpysqdfrevqsefodipfrgkyfwkwy 220  
QY 410 TEVQSGQRCENCRPRGRFYVVRHTEKVDGTLQCPAPDPCVAGCLSPGCCGILGSGR 469  
Db 221 -rgsgvkacsltclaeafnyteraavvdgtpcrpdtvdcvsgckhvgcdvrgsd 279  
QY 470 RPDGCGVGGDDSPCLVSGNLTRDGGPLGQKLTIPAGALRLQIAQLRPPSNYLALRG 529  
Db 280 redkrcvvggdsacetiegvfspagagdyvvpkpsvhiqdlnlslshlalkg 339  
QY 530 PGRSIIINGNAVDPPGYSRAGGTVFYRNPPREEKGESLSAEGPTTQPDVYMFQEE 589  
Db 340 dgesllleglpgtpqphrlplagttfqlrggpdv---qalealgpinaalvmvlarte 396  
QY 590 NPGVFYQY-----VSSPPPILENTPEPV-----POLQPEILRYE-PPLAPAPRPAR 637  
Db 397 lpalryrlnaplaridslppyswhypwtksaqcagsgqvqavecrnqlssavaphycs 456  
QY 638 TPGTL-QRQVRIPOMAPPPHPTPLGSPAAYWKRVGSHSACSACGKGVWRPIFLCISRES 696  
Db 457 ahsklprkraentepcpdd-----wvvgvswlcsrscdagvrsrvvcgrivs 505  
QY 697 GEE---LDERSCAAGARPPASPEFCHGTPCPPYWEAGETWSCRSQCGPTQHRQLQCRQE 753  
Db 506 aaeekalddscpg-prppvl--eachgtpcpwaaaldwsectpscgpglrhrvvlcks- 562  
QY 754 FGGGSSVPPERCGLHLPNPTTQSCQLRLC--GHWEVGSPPWSQCSVRCGRQVRVCV 811  
Db 563 -adrtatlpahcsapaakppatmrcnlrrcparrwage-wgecsaqcgvqgrsvrct 620  
QY 812 GNNGDEVSEQC 823  
Db 621 shtgq--ashec 630



```

424 QY 424 PRGRFVVRHTEKXVQDGTLCQGPAPDVCVAGRCLSPGCDGIILSGRRPRDGCVCVCGGDST 483
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
113 Db 113 pkgerfyirgrevdgtcrndkdldvcvngcempvgcdmmlgsgdakedckrkcgdgdst 172
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
484 QY 484 CRVSGMLTDRGCGPLGYQKILTPAGALRLQTAQLRSNYYLALRGGSRSIINGNAVD 543
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
173 Db 173 ckiirtittkdapayndllllpegatinrieeetvpsnylacnhsghyylngdwrid 232
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
544 QY 544 PPGSYRAGGTVFYRNPPEEGKGESLSAEGTTQPDVYMYFQENPGVFQYVISFPP 603
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
233 Db 233 fprpmffanswmyqrkpmgfaopdgtlcsqglseflfvlmvgknslsldyeysi---- 288
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
604 QY 604 PILENTPPEPVPQLOPEILRVEPPIAPAPRPARPTGTLQROVKRIPQMPAPPHPRTPLGS 663
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
289 Db 289 -----pe-----slshsqgqdt----- 301
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
664 QY 664 PAAYKRVGHSACSGKGWVRPFLCISRSSEGLDERSCAAGARPASPEPCHGTIPC 723
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
302 Db 302 -----wthqfnacsaecgsgsgskvctcnrrltlaevnpslcdqsk-pveeqagctgpc 356
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
724 QY 724 PPYWEAGMTSCRSRG-FGTQHRQLQCQREFGGGGSVPPE-----RCGLHPRENITQS 777
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
357 Db 357 aphwvegawsksgcgsgdgfnrsitcerisssgehtveedavclkevgn--kpatqge 414
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
778 QY 778 C--QLRLCGHWEVGSFWSQCSVRCGGRSQRVRC-VGNNGDE--VSQECASGPPQPS 832
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
415 Db 415 cndvknpcpyhlg-pwtcdklcgdgkdrkvctcfieenghkrvlpeedcve--ekpet 471
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
833 QY 833 REACDMGPCT-TAFWHSOWS 851
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||
472 Db 472 eksclltpceqvdiwlsqws 491
      : : ||| ||| ||| ||| : : ||| : : ||| : : ||| |||

```

**RESULT** 12

XX	AB58064	
ID	ABB58064	standard; Protein; 3060 AA.
XX	ABB58064;	
XX	26-MAR-2002	(first entry)
XX	Drosophila melanogaster	polypeptide SEQ ID NO 984.
XX	Drosophila;	developmental biology; cell signalling; insecticide;
KW	pharmaceutical.	
KW	Drosophila melanogaster.	
OS		
XX	WO200171042-A2.	
XX	27-SEP-2001.	
XX	23-MAR-2001;	2001WO-US09231.
XX	23-MAR-2000;	2000US-191637P.
PR	11-JUL-2000;	2000US-0614150.
PR		
XX	(PEKE )	PE CORP NY.
PA	Venter JC, Adams M, Li PWD,	Myers EW;
XX	WPI;	2001-656860/75.
XX	N-PSDB;	ABL02167.
XX		
XX	New isolated nucleic acid	detection reagent for detecting 1000 or more
PT	genes from Drosophila	and for elucidating cell signalling and cell-cell
PT	interactions -	
XX	Disclosure; SEQ ID NO 984;	21pp + Sequence Listing; English.
PS		
XX	The invention relates to an	isolated nucleic acid detection reagent
CC	capable of detecting 1000	or more genes from Drosophila. The invention is
CC		

CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 3060 AA;
	Query Match 14.5%; Score 711.5; DB 22; Length 3060;
	Best Local Similarity 31.4%; Pred. No. 4e-35;
	Matches 157; Conservative 78; Mismatches 190; Indels 75; Gaps 14;
QY	364 CSGSEQLRACSAQCPCPEQDPRALOCAFNSQEFMGQLYWQEPTEVQSGORCELNCR 423
DB	: :     :        :    :    :    :    :    :    :
DB	55 csgrsyfscntqdcpeesdraqgcsrfdlqqfdgvfyevvypnnp--celncm 112
QY	424 PRGRFVVRHTEKVDGTLQCGAPDLCVAGRLSPGCCGILGSRPDCGGVCGGDDST 483
DB	:                :    :    :    :    :    :    :
DB	113 pkgerfyrkdvkdgtcndkdlvcvngcempvcdmmldgakckcrkgdgst 172
QY	484 CRLVSGNLTRGGPLGYQKILWIPAGALRLQIAQLRPSSNYLALRGPGRSIINGWAYD 543
DB	: : :  :    :    :    :    :    :    :    :    :
DB	173 cktirnttkdlapgyndllllpegatnriceetvpssnylacrnshghyylngdwrid 232
QY	544 PPGSYRAGGVFRYRNPRREGKGESLASGPTQPVDVMIRIQENPGVFQYVTSPP 603
DB	: :  :    :    :    :    :    :    :    :    :
DB	233 fprpmffanswnyqrkpmgfaapdqitcsgpsieslfvmlygeknlsidyeys1---- 288
QY	604 PILENPPEPVPOLQEILRVPEPLAPAPARTPGLQRQVRIPQMAPPHPRTPLGS 663
DB	-----pe-----slshsqgdht----- 301
QY	664 PAAYWKRVGSHSACSACGKWGRPFIICISREGEELDERSCAAGARPAPSPFCHGTPC 723
DB	:    :    :    :    :    :    :    :
DB	302 ---wthhqfnaceascgggsqsrkvtnritlaevnpslcdqsk-pveeqacgtepc 356
QY	724 PPYEAGEWTSCSRSC-PGTQHRQLCROCFEGGGGSSVPPE-----RCGHLPKPNTQS 777
DB	:    :    :    :    :    :    :    :    :
DB	357 aphvwegewsksgcgdsdgnfrsitceerissseghtveedavclkevgn--kpatkqe 414
QY	778 C--OLRLCGHWEGSPWSQCSVRGGRGRQVRQC-VGNNGDE--VSEQECASGPQPPS 832
DB	: :  :    :    :    :    :    :    :    :    :
DB	415 cnrdvkncpkylhg-pwtpcdklgcdgkqrkrvtctfieenghrvriveedcve--ekpet 471
QY	833 REACDMGPCT-TAMFHSDWS 851
DB	:    :    :    :
DB	472 ekscllttpegvdlisqws 491
	RESULT 13
AA	AB72300
ID	AA72300 standard; Protein; 1072 AA.
XX	
AC	AA72300;
XX	
DT	14-MAY-2001 (first entry)
XX	
DE	Human ADAMTS-10 alternative amino acid sequence.
XX	
KW	ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
KW	tumour cachexia; inflammation;; dermatosparaxis; EDS-VIIC; angiogenesis;
KW	Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
KW	metastasis; embryogenesis; egg implantation; ADAMTS-10.
OS	Homo sapiens.
XX	
PN	WO200111074-A2.
XX	





Matches	185;	Conservative	66;	Mismatches	264;	Indels	76;	Gaps	19;
Qy	287	QGWASPVQVARRRPPDPFSPVRG-RGQQQ-GPWGTGGPHGRPLRPDQHPCAWLPLLSN	344	: :	:	:	:	:	:
Dd	505	kwcykrcv-----vpfgrpegvgawpwpgdcstcggvsssrhdcsprptigg	560	:	:	:	:	:	:
Qy	345	GPHASSLWSLFAPSSPIPRCSGESEQLRSCAQCPPEPQPDRALQCAAFNSOEFMGOLY	404	:	:	:	:	:	:
Dd	561	-----kycigerirrhscntddoppgsqdfrevqcseafdsiprfgky	603	:	:	:	:	:	:
Qy	405	QWEPFTEVOGSORCELNCRPGFRFYRHTEKVODTLCQCAPDICVAGRCGLSPGCCDI	464	:	:	:	:	:	:
Dd	604	kwkly-rggvkaacsitelaegfnfyteraavvdgtpcrpdtdicvsgeckhvgcdrv	662	:	:	:	:	:	:
Qy	465	LGSRRPDGCGVGCGDDSTCLRVSGNLNDRGGPLGYQKILWPAGALRLQAQLRPSNNY	524	:	:	:	:	:	:
Dd	663	lgsdlredkcrvcgdgsacetiegvfsgpsgagyedvvwipkgsvhifqdlnlslsh	722	:	:	:	:	:	:
Qy	525	LALRGPGGRSIIINGNAVDPPGSRAGCTVFRYNPREEKGESLSAEGPTQTQPDVVM	584	:	:	:	:	:	:
Dd	723	lalkgdesllleglpqtcpqhrlplagtftqlrgpgqv---qslsalgpinalimvw	779	:	:	:	:	:	:
Qy	595	IPOENPGVFYYQ----VISPPPILENTTPEPPV-----PQLQPIILRVE-PPLAPA	632	:	:	:	:	:	:
Dd	780	lartelpalryfnapiardslppyswhyapwtksacagsgqvavecnqlgdsava	839	:	:	:	:	:	:
Qy	633	PRPARTCTL-QRQVRIQMPAPHPRPPLGSPAAYWRVGHSAACSACSGKWGPRLC	691	:	:	:	:	:	:
Dd	840	phycsaahklpkkracnteppqd-----wvgnwslcsrdagvrstvvc	888	:	:	:	:	:	:
Qy	692	ISRSAGEE--LDERSCAAGARPPASPBPCHGTPCPPYWEAGETWCSCRSGPQTHROL	748	:	:	:	:	:	:
Dd	889	qrivsaaekalddsacpq-prppvI-eachgtpcpewatldwsecpscpgilhrvv	946	:	:	:	:	:	:
Qy	749	OQRPFEGGGSVPPERCGLHLPNPITQSCQLRLC--GHWGVGSPWSCSVRCRGQRSR	806	:	:	:	:	:	:
Dd	947	icks--adgratlbgblpaakpstnrcnlrrccparw-vtseggestcgigqqqr	1003	:	:	:	:	:	:
Qy	807	OVRVGNNGDEVSEQECASGPPQPPSREACD-MGPCTTAFWHSDMSKVSP	856	:	:	:	:	:	:
Dd	1004	tvtctshgq-----psrectealrpstmqceakcdsvpp	1040	:	:	:	:	:	:
RESULT	15								
AAB41379	ID	AAB41379 standard; Protein; 1784 AA.							
XX	XX								
AC	AAB41379;								
XX	XX								
DT	08-FEB-2001	(first entry)							
XX	XX								
DE	Human OREFX ORFL143	polypeptide sequence SEQ ID NO:2286.							
XX	Human;	open reading frame; OREFX; detection; cytostatic; hepatotropic;							
KW	vulnerary;	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;							
KW	anticongulsant;	osteopathic; antiarthritic; immunosuppressant; cardiac;							
KW	immunostimulant;	thrombolytic; coagulant; vasotropic; antidiabetic;							
KW	hypotensive;	dermatological; immunosuppressive; antiinflammatory;							
KW	antiviral;	antibacterial; antifungal; antihematic; antithyroid;							
KW	antiamebic;	gene therapy; cancer; proliferative disorder; hypertension;							
KW	neurodegenerative disorder;	osteoarthritis; graft vs host disease;							
KW	cardiovascular disease;	diabetes mellitus; hypothyroidism; SCID; AIDS;							
KW	cholesterol ester storage;	systemic lupus erythematosus; infection;							
KW	severe combined immunodeficiency;	malaria; autoimmune disorder; asthma;							
KW	allergy; aplastic anaemia;	nocturnal haemoglobinuria; burn; wound;							
KW	bone damage; cartilage damage;	antiinflammatory disease; coagulation;							
XX	thrombosis; contraceptive.								
OS	Homo sapiens.								
XX	XX								
PN	W0200058473-A2.								
XD	05-OCT-2000.								

Qy	650	QMPAPPH-----PRTPL-----GSPAAVWKRV-----GH--	673
		:	
Db	856	nivaavhrggwaqplggvrrhlvimgprlptqllfgesnbgvhyetihr eagghde	915
		:	
Qy	674	-----SACSACGKGYWRPFLGISRESGEELDERSCAAGARPPASPpPCHG	720
		:	
Db	916	vpppvfswhygpgwtKctvtcgrgvqrqvyelerqag-pvdeehcdplgprddqgkcs	974
		:	
Qy	721	TPCpPYWAGEWTSGRSGGP-GTQHROLOCRQBFQ-GGSSVPPPERCGHLPRPNTQSC	778
		:	
Db	975	qpcparwageqlcsscgpggisrtavlgrsvgidegsaieppaceahlpprtetpc	1034
		:	
Qy	779	QLRL-C-GHWEYSPWSOCYSVRCRGGORSRQVRCVGNNGDEVSEQECASGPPpPSREAC	836
		:	
Db	1035	nrhvpcpatwavn-wsqscvtcsegtgrnvlctndtgvpcdeaq-----qpasevtc	1087
		:	
Qy	837	DMGPC	841
		:	
Db	1088	sipic	1092

Search completed: July 23, 2002, 21:04:40  
Job time: 3729 sec

OM of: US-10-041-770-2 to: N\_Geneseq\_032802.\* out\_format : pfs  
 Date: Jul 23, 2002 10:33 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=framet-p2n.model -DEV=xlh  
 -Q/cgn2\_1/USF10.spool/US10041770/runat\_22072002.153136.2727/app\_query.fasta\_1.947  
 -DB=N\_Geneseq\_032802 -CFMT=fastap -SUFFIX=ring -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US10041770 -CGN1\_1\_206 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-10-041-770-2  
 Query length: 877  
 Database: N\_Geneseq\_032802.\*  
 Database sequences: 1736436  
 Database length: 85845721  
 Search time (sec): 240.610000

## score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	1.7e-6	1235.95	2371.00	1.7e-6	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	2.5e-1	925.00	496.89	2.5e-1	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	925.00	496.89	2.5e-1	925.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	808.50	428.09	1.7e-1	808.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	807.50	427.87	1.7e-1	807.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	806.50	427.66	1.9e-1	806.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	804.50	426.04	2.2e-1	804.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	797.50	424.99	2.5e-1	797.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	796.50	424.47	2.7e-1	796.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	796.00	423.51	3.0e-1	796.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	790.50	420.27	4.6e-1	790.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	789.50	421.46	4.0e-1	789.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	788.00	419.94	4.8e-1	788.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	780.00	416.86	7.1e-1	780.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	757.50	400.80	5.1e-1	757.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	753.50	401.79	4.9e-1	753.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	728.00	389.30	2.5e-1	728.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	714.50	384.29	4.7e-1	714.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	714.00	376.51	1.3e-1	714.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	699.50	364.62	5.8e-1	699.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	698.50	374.08	1.7e-1	698.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	676.00	361.36	8.9e-1	676.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	636.00	342.51	9.9e-1	636.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	626.50	339.70	1.4e-1	626.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	625.00	337.26	1.9e-1	625.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	618.00	328.88	5.7e-1	618.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	596.00	320.79	1.6e-1	596.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	590.50	317.12	2.6e-1	590.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	589.50	314.18	3.7e-1	589.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	589.00	314.68	3.5e-1	589.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	588.00	314.00	3.8e-1	588.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	587.00	315.58	3.1e-1	587.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	581.00	313.17	4.3e-1	581.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	581.00	310.57	6.0e-1	581.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	572.00	307.27	9.1e-1	572.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	566.00	304.11	1.4e-1	566.00	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	556.50	298.04	3.0e-1	556.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	556.50	297.99	3.0e-1	556.50	
/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 +	551.00	301.86	1.8e-1	551.00	

/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 + 551.00 298.40 2.8  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 + 547.50 297.09 3.4e  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 + 545.50 296.40 3.7e  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 + 541.00 279.36 3.3e  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 + 541.00 279.36 3.3

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048

seq\_documentation\_block:  
 ID AAC91178 standard; DNA; 2680 BP.

AC AAC91178;  
 DT 20-MAR-2001 (first entry)

DE ADAM gene #2.

KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;  
 KW ovarian; SS.

OS Homo sapiens.

PN WC200073323-A2.

PD 07-DEC-2000.

PF 25-MAY-2000; 2000WO-US14308.

PR 27-MAY-1999; 99US-0136388.

PR 09-JUL-1999; 99US-0142930.

PR 28-JAN-2000; 2000US-0178717.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;

WPI; 2001-016507/02.

Seven nucleic acid molecules encoding ADAM polypeptides containing a  
 disintegrin and metalloprotease domain, useful in the prevention,  
 treatment and diagnosis of cancer, immune disorders, cardiovascular  
 disorders and neurological diseases -

XX Claim 1; Page 266; 287pp; English.

XX The present invention relates to seven members of the ADAMs (proteins  
 which contain a Disintegrin And Metalloprotease domain) protein family.  
 The ADAMs proteins and DNA may be used to treat disease, as a food  
 additive or preservative, for chromosome identification, as probes  
 for diagnosing a disorder related to the female reproductive system,  
 particularly breast and/or ovarian cancer. They are also useful in the  
 gene therapy of breast and ovarian cancer.

XX Sequence 2680 BP; 510 A; 864 C; 767 G; 537 T; 2 other;

alignment\_scores:  
 Quality: 2371.00 Length: 423  
 Ratio: 5.632 Gaps: 0  
 Percent Similarity: 99.527 Percent Identity: 99.527

alignment\_block:

US-10-041-770-2 x AAC91178 ..

Align seg 1/1 to: AAC91178 from: 1 to: 2680

435 GILUYSVALGlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCy 451

5 GAAAGGTCAGAGGAGGAGCCCTGTGTGAGCCCTGAGCCCTGACACTG 54

451 sValaLaGlyArgCysLeuSerProGlyCysAspGlyLeuGlySerG 468

55 TGTGCTGAGCGCTGTCTGTGAGCCCGGCTGTGTGATGGGATCCTTGGCTCTG 104

468	lyArgArgProAspGlyIcYcSGlyValcYcSGlyGlyAspAspSerThrCys	484
105	GCAGGCGTCTGTGATGCGTGTGGAGTCTGTGGGGGTGATGATTTCACTGCT	154
485	ArgLeuValSerGlyAsnLeuThrAspArgGlyVcGlyProLeuGlyTyrgI	501
155	CGCTTTGTTTGGGGAACCTCACTACCGAGGGGGGCCCCGTGGCGTATCA	204
501	nlystIleLeuTrrPileProAlaGlyAlaLeuArgLeuGlnIleAlaGlnL	518
205	GAAGATCTTGTGGATTCAGCGGGAGCCCTGCGGCTCCAGATTGCCAGC	254
518	euArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSer	534
255	TGCGGCGTGTAGCTCAACTACTGTGCGACTTGTGTGCGCTGGGGCGCGTCC	304
535	IleIleAsnGlyAsnTrpAlaValAspProProGlyIserTyrArgAlaGln	551
305	ATCATCAATGGGAACCTGGGCTGTGGATCCCCCTGGGTCTACAGGCCCG	354
551	YglyThrValPheArgTyrAsnArgProProArgGluGluGlyLysGlyG	568
355	CGGGACCGTCTTTTCGATATAAACGTCCTCCACAGGGAGAGGGCAAGGGG	404
568	luserLeuSerAlaGluGlyProThrThrGlnProValAspValTyrMet	584
405	AGAGTCTGTGGCTGAAGGCCCCACCACCCAGCGCTGTGGATGTCTATATG	454
585	IlePheGlnGluGluAsnProGlyValPheTyrGlnTyrValIleSerse	601
455	ATCTTTTCAGGAGGAAAACCCAGCGCTTTTATCAGTATGTCATCTCTTC	504
601	rProProProIleLeuGluAsnProThrProGluProProValProGlnL	618
505	ACCTCTCTCAATCCTGTAGAACCCACCACCCAGAGCCCCCTGTCCCCAGC	554
618	euGlnProGluIleLeuArgValGluProProLeuAlaProAlaProArg	634
555	TTCAGCGCGAGATCTCAGGGTGGAGCCCCCACTTGTCTCGGCACCCCGC	604
635	ProAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGlnMe	651
605	CCAGCCCGGAGCCCGAGCACCTTCCAGCGTCAGTGGCGATCCGCCAGAT	654
651	tProAlaProProHisProArgThrProLeuGlySerProAlaAlaTyrT	668
655	GCCTGCCCCCCCCATCCACAGACACCCCTGGGGTCTCAGCTGCGTACT	704
668	rpLysArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyVal	684
705	GGAAACGAGTGGGACACTTCGATGCTCAGCGTCTCGGGAAAGGTGTC	754
685	TrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAs	701
755	TGCGGCCCATTTTCTCTGCATCTCCGHTAGTCTGGGAGAGGAACCTGGA	804
701	pGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluProC	718
805	TGAACGCACTGTGCGCGGGTGCAGGCCCCAGCGCTCCCTCGAACCT	854
718	ysHisGlyThrProCysProProTyrTrpGluAlaGlyGluTrrThrSer	734
855	GCCACGGCACCCCATGCCCCCATCTCTGGGAGGCTGGCGAGTGGACATCC	904
735	CysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysAr	751
905	TGCAGCGCTCCTGTGGCCCCGCGCACCCACGACCGCCAGCTGCAGTCCG	954
751	gGlnGluPheGlyGlyGlySerSerValProProGluArgCysGlyH	768
955	GCAAGAAATTGGGGGGGTGGCTCTCTCGTGTGCCCGCCGAGCGCTGTGGAC	1004

768	isLeuProArGProAsnIleThrGlnSerCysGlnLeuArgLeuCysGly	784
1005	ATCTCCCGCGCCCAACATCACCCAGTCTTGCCAGCTCGCGCTCTGTGGC	1054
785	HisTrpGluValGlySerProTrpSerGlnCysSerValArgCysGlyAr	801
1055	CATTGGGAAGTTGGCTTCCTTGGAGCCAGTGTCCGTGCGGTGGGCGG	1104
801	gGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluV	818
1105	GGGCCAGAGAACCGCGCAGGTTGCTGTGTGGGAACAACCGTGATGAAG	1154
818	alSerGluGlnGluCysAlaSerGlyProProGlnProProSerArgGlu	834
1155	TGAGCGACGAGAGTGGCTGAGCGCCCGCCACAGCCCCCGCCAGCAGAG	1204
835	AlaCysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTrpSe	851
1205	GCCTGTGACATGGGCGCCTGTACTACTGCTGTGTCCACACGCGACTGGAG	1254
851	rSerLysValSerProGlu	857
1255	CTCCAAGTGCTCAGCCGAG	1273
seq_name: /SIBS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAD16773		
seq_documentation_block:		
ID:	AAD16773 standard; cDNA; 1020 BP.	
XX		
AC	AAD16773;	
XX		
DT	22-NOV-2001 (first entry)	
XX		
DE	Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:34.	
XX		
KW	Human; cytostatic; gene therapy; inflammatory disorder; neural disorder	
KW	Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS	
KW	autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemic	
KW	reproductive disorder; Crohn's disease; pulmonary disorder; cancer;	
KW	myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour	
KW	haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;	
KW	anti-sense therapy; endocrine disorder; leukaemia; ss.	
OS	Homo sapiens.	
XX		
FH	Location/Qualifiers	
FT	65..847	
FT	/tag= a	
FT	/product= "Human novel protein"	
FT	/trans_except= (pos:767..769, aa:Xaa)	
FT	/note= "CDS does not include start codon; Xaa is an	
FT	unknown amino acid"	
FT	/partial	
XX		
PN	WO200155202-A1.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01325.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209457.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	



cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias) CC  
renal disorders (glomerulonephritis, nephrotic syndrome); cancerous CC  
disease and conditions (breast cancer); hyperproliferative disorders CC  
(leukaemias, myelomas); tumours: foetal and developmental CC  
anomalities; haematopoietic disorders; respiratory disorders (rhinitis, CC  
asthma); angiotonic disorders; diabetes; atherosclerosis; endocrine CC  
disorders; pregnancy-related disorders and infections. The novel protein CC  
DNA is useful in gene therapy and anti-sense therapy. The proteins can

alignment_scores:		
Quality:	925.00	Length: 172
Ratio:	5.572	Gaps: 0
Percent Similarity:	96.512	Percent Identity: 96.512

```
alignment_block:      x AAD16773
```

Align seq 1/1 to: AAD16773 from: 1 to: 1020

686	ArgProIlePheLeuCysIleSerArgGluSerGlyGluLeuAspG1	702
15	CGTCCGCCACGGCTCCGATCTCCGTGAGTCGGAGAGAACTGGATGA	64
702	uArgSerCysAlaLaGlyValaArgProProAlaSerProGluProCysH	719
65	ACGCAGCTGTCCCGGGGTGCCAGGCC. CCAGCGCTCCCTGAAACCTGCC	113
719	IsGlyThrProCysProProTyrrTrpGluAlaGlyGluThrThrSerCys	735
114	ACGGCACCCCATGCCCCCCATCTGGGAGGCTGGCAGTGGACATCTCTGC	163
736	SerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgG1	752
164	AGCCGCTCTCTGGCGCCCGGCACCCAGCACCGCCAGCTGCAGTGCCTGGCA	213
752	nGluPheGlyGlyGlyGlySerSerValProProGluuArgCysGlyHisL	769
214	GGAAATTGGGGGGGTGGCTCCTCGGTGCCCGCCGACGGCTGTGGACATC	263
769	euProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHisL	785
264	TCGCCCGGCCCAACATCACCCAGATCTGCCAGCTGCGCCCTCTGTGGCCAT	313
786	TrpGluValGlySerProTrpSerGlnCysSerValArgCysGlyArgG1	802
314	TGGGAAGTTGGCTCTCCTTGGAGCCAGTCTCGCTGCGGTGCGGCCGGGG	363
802	YGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValS	819
364	CCAGAAAGCCCGCAGGTTCCGTGTGGGAACAATGGGAGTGAAGATGA	413
819	erGluGlnGluCysAlaSerGlyProProGlnProProSerArgGluAla	835
414	GCAGCAGGAGTGTGGCTCAGCGCCGCCACAGCCGCCCCAGCAGAGAGGCC	463
836	CysAspMetClyProCysThrThrAlaTrpPheHisSerAspTrpSerSe	852
464	TGTGACATGGGGCCCTGTACTACTCTGCTGGTTCCACAGCAGCTGGAGCTC	513
852	rLysValSerProGlu	857
514	CAAGTGTCTACGGCAG	529

seq name: /stps1/qcadata/hold-geneseq/geneseq-emb1/NA2001A DAT: AAD16758

seq documentation block:

seq\_documentation\_block:  
ID AAD16758 standard: cDNA: 1998 BP:

XX AAD16758:

22 XX

DT 22-NOV-2001 (first entry)

Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:19.

2

XX	Human; cytostatic; gene therapy; inflammatory disorder; neoplasia
KW	Parkinson's disease; Alzheimer's disease; immune system disorders
KW	autoimmune disease; rheumatoid arthritis; muscular disorders
KW	reproductive disorder; Crohn's disease; pulmonary disorders
KW	myocardial infarction; glomerulonephritis; nephrotic syndrome
KW	haematopoietic disorder; rhinitis; asthma; diabetes; arthritis
KW	anti-sense therapy; endocrine disorder; leukaemia; ss.
XX	
XX	Homio sapiens.
XX	
FX	Location/Qualifiers
FX	56..1174
FT	/*tag= a
FT	/product= "Human novel protein"
FT	/note= "CDS does not include start codon"
FT	/partial
FT	

```
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234597.
PR 23-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451925/48.
XX P-PSDB; AAE09698.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX medical disorders and also for testing and detection e.g. diagnosis and
XX screening for agonists -
XX
XX Claim 1; SEQ ID NO 19; 469pp; English.
XX
XX AAD16750-AAD16775 represent cDNAs corresponding to novel human protein
XX genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777
XX -AAD16780 represent novel human genomic DNA fragments. The novel proteins
XX and their DNAs are useful for diagnosing, treating, preventing and/or
XX prognosing inflammatory disorders (bursitis or tendonitis); neural
XX disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system
XX disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);
XX muscular disorders; reproductive disorders; gastrointestinal disorders
XX (malabsorption syndrome, Crohn's disease); pulmonary disorders;
XX cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);
XX renal disorders (glomerulonephritis, nephrotic syndrome); cancerous
XX disease and conditions (breast cancer); foetal and developmental
XX (leukaemia, hyperplasia); tumours; foetal and developmental
XX abnormalities; haematopoietic disorders; diabetes; atherosclerosis;
XX asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine
XX disorders; pregnancy-related disorders and infections. The novel protein
XX DNA is useful in gene therapy and anti-sense therapy. The proteins can
XX also be used to aid wound healing and epithelial cell proliferation,
XX to prevent skin aging due to sunburn, to maintain organs before

alignment_scores:
    Quality: 925.00      Length: 172
    Ratio: 5.572        Gaps: 0
    Percent Similarity: 96.512    Percent Identity: 96.512

alignment_block:
US-10-041-770-2 x AAD16758 ..
Align seg 1/1 to: AAD16758 from: 1 to: 1998

686 ArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAspG1 702
|||||
6 CQTCCGCCACCGCTCCGATCTCCGTCGAGCGGAGAGAACTGGATGA 55
|||||

702 uArgSerCysAlaAlaGlyAlaArgProAlaSerProGluProCysH 719
|||||
56 AGCAGAGTGTGCGCGGTGCCAGGCC.CCAGCCTCCCTGAACCTGCC 104
|||||

719 IsGlyThrProCysProProTyrTrpGluAlaGlyGluTrpThrSerCys 735
|||||
105 ACGGCACCCCATGCCCCCATCTGGAGGTGGCGGAGTGCATCCTGC 154
|||||

736 SerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgG1 752
|||||
155 AGCCGCTCTGTGGCCCGGCCACCCAGCACCCAGCTGCAGTGCCTGCCGA 204
|||||
```





```

881 CATTGGGAACCCATCCACATCCATTGTGGCTGCTGCTGCTGGAAG 930
      :::::||||:
66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro..... 80
      ||| ||||| ::| ||::||| |||
931 ATGAGGAGGAGCACTAAGATTCAGCCACCATGCAAGCAACACCCCTGAAG 980
      ||| ||||| ::| ||::||| |||
81 .....SerLeuProLeuP 85
      :::::||||:
981 AGCTTCTCAAGTGCAGAAAGCATCAACATGAAGGGGATGCCATCC 1030
85 roProArgProProArgHisProGluAlaLeuProArgGlyGlnGly 101
      ||
1031 COTGCACCATG..... 1041
102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118
      ||::: ||||| |||
1042 ACACCTGCCATCTGCTCACAGAA.....AGCACCTGTGT 1076
118 rArgGlyArgGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
      ::| |||||::|||
1077 GCAGCCATGAACCGCCCTGTGAGACCCCTGGACTGTCCCATGTGGCGGG 1126
135 luGluThrGlnGluIleArgAlaAlaArgSerArgLeuArgAspPro 151
      :::::||||:
1127 CATGTGCCAGCGCACCGAGCTGCAGCATCAACAGGACA..... 1167
152 IleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuProLe 168
1167 ..... 1167
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
      ||| ||| ||| ||| ||| |||
1168 ...CGGGCTCGCGTGCCTTCACTAGCCACGAGCTCGGGCACAG 1213
185 euLysSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201
      :::::||||:
1214 TTTTGCATTCAGCATGACGAGCGGCAATGACTGTGAGCCCGTTGGGA 1263
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
      ::| ||||| |||
1264 AGCAGCTTTCATCATGTCTCCACAGCTCCTGTACGACCCGCTCCCTC 1313
216 .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
      ||| ::::: ||| ||| |||
1314 ACCTGTCCCGTGCAGCGCCAGTATATCACAGGTTCCTTGACCGTGG 1363
232 exProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
      ::| ||| ||| ||| ||| |||
1364 GTGGGCGCTGTGCTGGAGACCCCTCCTGCCAAGGACATTATCGACTTCC 1413
249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258
      ||| |||
1414 CCTCGTGCCACCTGGCGTCTCTATGATGTAAGCCACCATGTCGCCCTC 1463
258 rGlyThrGluProProSerPro.....T 266
      ||::||| ||| |||
1464 CAGTACGGGGCGCTACTCTCCCTCTCCGAGGACATGATTAATGTGCCA 1513
266 hrHisSerLeuGlyGluGlyPhePheArgAlaSerProGlnProArg 282
      ||||| ||| ||| |||
1514 CACACTCTGGTCTGTGGGA.....CCACTGTCT 1545
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg... 297
      ||||| ||| ||:::||||| |||
1546 ACTCCAAGCTGATGACGCGGTGGAGCGGACCGCGGTGTG...GGGAGAAT 1592
298 .....ArgProAspProPheProSerValProArg 308
      ||||| ||| |||
1593 AGTGTGTCTCAGTGGGAGTGCCTACCGTGGGCTTCGGGCCGAGG. 1641
308 lYargGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321
      ||||| ||| ::|||

```

```

1642 .....CCGTGGATGGTGGCTGCTGCTGCTGGA 1668
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338
      ::||| |||||
1669 GCGCTT...GGTCCA..... 1680
338 pLeuProLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355
      ::| ||:::||||:
1681 .....TCTGCTCAGGAGTGTGGCATGGCGGTACA 1711
355 heAlaProSer...Ser.....ProIleProArg..... 363
      ||||| ||| ::|
1712 GAGCGCGGAGCGGAGTGCAGCGAGCTACGCCCAATACAAAGGCAGAT 1761
364 ..CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCys 379
      ||| ||||| :::::||||| |||
1762 ACTGTGTGGTGGAGCGCAAGCGTTCGCTCTGCAACCTGCAGGCCCTGC 1811
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSe 396
      ||| :::::||||:
1812 CCTGCTGGCCGCCCTCTCTCCGCCAGCTCCAGTGCAGCCACTTGGACG 1861
396 rGlnGluPheMetGlyGlnLeuTyArgGlnProPheThrGluValG 413
      ::| ::| |||||::||| |||
1862 TATGCTCTACAAGGCCAGCTGCACACATGGTCCCGTG.....GTCA 1905
413 lnGlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhe 429
      ::| ::| |||||::||| |||
1906 ATGACGTGAACCCCTGGAGCTGCATGCCGCCCGCGAATGAGTACTTT 1955
430 TyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro... 445
      ::| ::| ||| ||||| |||
1956 GCCGAGAAGCTCGCGGACGCGTGTGTCATGGCACCCCTGCTACCAGGT 2005
446 ....GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyC 461
      ::::: |||||::||| ||| |||
2006 CCGAGCCAGCGGAGCTCTCATCAAGCATCTGTAAGAACGCTGGCT 2055
461 ysAspGlyIleLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477
      ||||| ::| ||||| |||
2056 GTGACTTCGAGATTCACTCCGCTGCTATGGAGGACCGCTGTGGTGTGTC 2105
478 GlyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspAr 494
      ||::: |||||::||| |||||::|||
2106 CACGGCAACGGCTCCACCTGCCACCGTGGCGGACCTTCGAGGAGGC 2155
494 gGlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaL 511
      ||| ||||| |||
2156 CGAGGC...CTGGGTATGTGGATGTGGGCTGATCCACGCCGCGCAC 2202
511 euArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeu 527
      :::::||||:
2203 GCGAGATCCGATCCAGAGGTTCGCGAGGTGCACCTTCACATACGACGC. 2252
528 ArgGlyProGly...GlyArgSerIleIleAsnGlyAsnTrpAlaValas 543
      ||||| ::| ||||| |||
2253 CGAGTGAAGGACCGGAGAGTACTTCTCAATGGTGGCTGGACCATCCA 2302
543 pProProGlySerTyArgAlaGlyGlyThrValPheArgTyrAsnArgp 560
      ||::: |||||::||| ||| |||
2303 GTGGAACGGGACTACCAAGTGGCAGGACCACTTCACATACGACGC. 2351
560 roProArgGluGluGlyLysGlyGluSerLeuSerAlaGluGlyProThr 576
      ::| |||||::||| |||||
2352 .....AGGGCAACTGGGAGAACCTCAGTCCCCCGGTCCACC 2390
577 ThrGlnProValAspValTyMetIlePheGlnGluGlnAsnProGlyVa 593
      ||||| ::| :::::|||||
2391 AAGGAGCTGTCTGGATCCAGTGTGTTCAGGAGAGCAACCTCGGGGT 2440
593 lPheTyArgGlnTyValIleSerSerProProIleLeuGluAsnProT 610
      ||| |||||
2441 GCACCTACGAGTAC..... 2453

```

```
610 hrProGluProValProGlnLeuGlnProGluLeuArgValGlu 626
2453 .....
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG1 643
2454 .....ACCATCCA 2461
643 nArgGlnValArgIleProGlnMetProAlaProProHisProArgThrP 660
2462 CAGGGAG.....GCAGGTGGCCACGACGAGGTCC 2490
660 rLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCys 676
2491 CG.....CCGCCGTGTTCTCTGGCATATTGGCCCTGGACCAAGTGC 2534
677 SerAlaSerCysGlyGlyValTrpArgProIlePheLeuCysIleLe 693
2535 ACAGTCACCTGGCGCAGAGGTGTGCAGAGGAGCAATGTGTACTGCTTGG 2584
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAla 710
2585 CGCGCAGCAGG...CCGTGGCAGGAGGACACTGTGACCCCTGGGCC 2631
710 rgProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726
2632 GGCCTGATGACCAACAGAGGAAGTGCACGCGACGACCTGCCCTGCCAG 2681
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...G1 742
2682 TGGTGGCAGGTGAGTGGCAGCTGTGCTCCAGCTCTCGGGCCCTGGGG 2731
742 YThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGly 758
2732 CCTCTCCCGCGCGCGCTGTGTGATCCGACGCTGGGCTGGATGAGC 2781
758 LysSerSerValProGluArgCysGlyHisLeuProArgProAsnIle 774
2782 AGAGCGCCTGGAGCCACCGCCTGTGAACACCTTCCCGCCCTCTACT 2831
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValG1 789
2832 GAAACCCCTTGAACCCGCGCATGTACCTGTCCGCCACCTGGCTGTGG 2881
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSera 806
2882 GAAC...TGGTCTCAGTCTGATGACATGTGGGAGGGGACTCAGCC 2928
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822
2929 GAAATGTCTCTGCACCAATGACACCGGTGTCCCTGTGACGAGGCCAG 2978
823 CysAlaSerGlyProProGluProProSerArgGluAlaCysAspMetG1 839
2979 .....CAGCCAGCCAGGAATGACCTGTCTCTGCC 3010
839 yProCysThr.ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854
3011 ACTCTGTGGTGGCCCTGGGCACAC.....TGGCCCTCAAGGCTCAG 3054
855 SerProGluProProAlaIleSer 862
3055 GCAGCGGTCTCTCCAGCCAGCAGC 3078
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF82149

seq\_documentation\_block:

ID AAF82149 standard; cdna; 5061 BP.

XX AAF82149;

AC

XX

DT 02-JUL-2001 (first entry)

```
XX Human ADAM type metal protease MDTs1 encoding cDNA SEQ ID NO:2.
XX
XX Human; a disintegrin and metalloprotease type metal protease; MDTs1;
KW MDTs2; MDTs3; ADAM type metal protease; cytosolic; antiarthritic;
KW cancer; arthritis; arthrosis deformans; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..5061
FT /tag= a
FT /product= "MDTs1"
FT /note= "a disintegrin and metalloprotease (ADAM)
FT type metal protease"
XX
XX JP2001008687-A.
XX
XX 16-JAN-2001.
XX
XX 25-JUN-1999; 99JP-0180973.
XX
XX 25-JUN-1999; 99JP-0180973.
XX
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
XX WPI; 2001-285362/30.
XX
XX P-PSDB; AAF82149.
XX
XX New metal protease and metal protease gene, for use as a drug for
XX cancers, arthritis and arthrosis deformans
XX
XX Example 1; Page 17-19; 31pp; Japanese.
XX
XX The present sequence encodes a disintegrin and metalloprotease (ADAM)
XX type metal protease designated MDTs1, isolated from human. MDTs proteins
XX have cytosolic and antiarthritic activities. They can be used as a drug
XX for cancers, arthritis and arthrosis deformans.
XX
XX Sequence 5061 BP; 878 A; 1746 C; 1587 G; 850 T; 0 other;
```

## alignment\_scores:

Quality: 807.50 Length: 959  
Ratio: 1.815 Gaps: 41  
Percent Similarity: 46.403 Percent Identity: 28.050

## alignment\_block:

US-10-041-770-2 x AAF82149 ..

Align seg 1/1 to: AAF82149 from: 1 to: 5061

```
15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24
596 GTGCTCCACGACCTGTGGAGTGAAGTGTACCCAGAGCTGGAGTCTCGA 645
24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
646 CGGAGCGTTGGGAGCAGCGGACGAGTGGCGCGCCAGCGCTGAGCG 695
35 hrProThrGluGluGlyGlnGlyProGluGlyValTrpGlyProTrp... 50
696 TCTACACCGGTGCTGTCAGCAAGAGAGAGTGGTGGAGACCTGTGTAG 745
50 ..... 50
746 TAGCTGATGCCAAATGGTGGAGTACCACGACGCGCAGGTGAGAGC 795
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
796 TATGTGCTGACCATCATGAATGGTGGCTGGCTTTTCATGACCCAG 845
61 lyValGlyVal.....Gln 65
```

```

846 CATTTGGAAACCCATCCACATCACCATTGTGGCGCTGGTCTGCTGGTGAAG 895
      ::::|||||:
66  ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80
      ||| ||||| :::: ||||| |||
896 ATGAGGAGGAGGACCTAAGATCAGCCACCTGCGAGACACACCTGAAG 945
81 .....SerLeuProLeuP 85
      ::::|||||:
946 AGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCCCATCC 995
85 roProArgProProArgHisProGluAlaLeuProArgGlyGlnGly 101
      ||
996 CTGTCACCATG..... 1006
102 ProArgProGlnThrSerProGluThrLeuProLeuTyrArgThrGlnSe 118
      ||||: ||||| |||||
1007 ACACCTGCCATCTGCTCCACAGAA.....AGGACCTGTGT 1041
118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
      :::: |||||: |||||
1042 GCAGCATGAAACCGCCCTGTGTGAGCCCTGGGACTGTCCCATGTGGCGG 1091
135 luGluThrGlnGluLeuArgAlaAlaArgSerArgSerArgLeuArgAspPro 151
      :::: ||||| :|||
1092 CATGTGCCAGCGCACGCGAGCTGCAGCATCAACGAGGACA..... 1132
152 IleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuProLe 168
1132 ..... 1132
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
      ||| ||| ||| ||||| ::::
1133 ....CGGGCTCGCGTGGCCCTTCACTGTAGCCACGAGCTGGGCACAG 1178
185 euLleSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201
1179 TTTTGGCATTCAGCATGACGAGCGGCAATGACTGTGAGCCCGTTGGGA 1228
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
      :::: |||||: ||||| |||||
1229 AAGCACTTTTCATCATGCTCCACAGCTCTGTACGACGCGCTCCCTC 1278
216 .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
      ||| :::: ||||| :|||
1279 ACCTGTCTCCGCTGCAGCGCGCATATACACAGGTTCTCTTGACCGTGG 1328
232 exProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
      :||| ||| ||| ||||| |||||
1329 GTGGGCGCTGTGCTGGAGCAGCCCTCTGTCGCAAGACATATTCGACTTCC 1378
249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258
      ||| ||| :|||
1379 CCTCGTGCACCTGGCGPCCTCTATGATGTAAGCCACGAGTGGCGCTC 1428
258 rGlyThrGluProProSerPro.....T 266
      ||||| ||| |||
1429 CAGTAGGGGCCCTACTCTGCTCTGCGAGGACATGGATATGTGTGCCA 1478
266 hrHisSerLeuGlyGluGlyPhePheArgAlaSerProGlnProArg 282
      ||||| :||| |||
1479 CACACTCTGTGCTCTGTGGGA.....CCACCTGTC 1510
283 ArgProSer...SerGlnThrPalaSerProGlnValAlaGlyArg... 297
      ||||| ||| ||||| |||||
1511 ACTCCAAGCTGGATGACGCGGTGAGCGGACCCCGTGTG...GGGGAAT 1557
298 .....ArgProAspProPheProSerValProArgG 308
1558 AAGTGGTGTCTCAGTGGGAGTGCGTACCCGTGGGCTTCGGGCCCGAGG. 1606
308 lyArgGlyGlnGlnGlyGlnGlyProThrP.....GlyThrGlyGly 321
      ||||| ||| :|||
```

```

610 hrProGluProValProGlnLeuGlnProGluIleLeuArgValGlu 626
2418 .....
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlu 643
2419 .....ACCATCCA 2426
643 nArgGlnValArgIleProGlnMetProAlaProProHisProArgThrP 660
2427 CAGGGAG .....GCAGGTGGCCACGACGAGGTCC 2455
660 LeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCys 676
2456 CG.....CCGCCCTGTCTCCTCGCATATATGGCCCTGGACCAAGTGC 2499
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693
2500 ACAGTCACCTGGCCAGAGGTGTGCAGAGGCAGAAATGTACTGTGTGA 2549
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710
2550 GCGGAGGCGAGG...CCGCTGGAGGAGGACACTGTGACCCCTGGGCC 2596
710 rGProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726
2597 GSCCTGATGACCAACAGAGGAAGTGCAGCGACGACCTGCTGCCCTCCAGG 2646
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...G 742
2647 TGGTGGCAGGTGAGTGGCAGGTGTGCTCCAGCTCTGCGGCGCTGGGG 2696
742 YthrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGly 758
2697 CCTCTCCCGCGGGCGTCTCTGTCATCGCAGCGTGGGTGGATGAGC 2746
758 LysSerValProGluArgCysGlyHisLeuProArgProAsnIle 774
2747 AGAGCGCCCTGGAGCCACCGCGCTGTGAACACCTTCCCGCGCCCTACT 2796
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValG 789
2797 GAACCCCTTGCACCGCATGACCTGCTCCGCCACCTGGGCTGTGGG 2846
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerA 806
2847 GAAC...TGGTCTCAGTCTCAGTGACATGTGGGAGGCGCACTCAGCGCC 2893
806 rGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822
2894 GAAATGTCTCTGCACCAATGACCGGTGTCCCTGTGACGAGGCCAG 2943
823 CysAlaSerGlyProGlnProProSerArgGluAlaCysAspMetG 839
2944 .....CAGCCAGCCAGGAGTCACTGCTCTCTGCC 2975
839 yProCysThr...ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854
2976 ACTCTGTCTGGTGGCCCTGGGCACAC.....TGGGCCCTGAAGGCTCAG 3019
855 SerProGluProProAlaIleSer 862
3020 GCAGCGGCTCTCTCCAGCCACGAGC 3043

```

seq\_name: /SID1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA04475

seq\_documentation\_block:

ID\_AA04475 standard; cDNA; 5353 BP.

XX

AC AD04475;

XX

DT 04-JUL-2001 (first entry)

```

XX Human 27875 ADAM-TS cDNA, alternative version.
DE
XX Human: ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW cardiac; tumour; thymoma; vasotropic; cytostatic; virucide; ss.
XX Homo sapiens.
OS
XX
XX Location/Qualifiers
FT Key 36..5096
FT CDS /tag= a
FT /product= "Human 27875 ADAM-TS protein"
FT /note= "Xaa is an unknown amino acid"
FT sig_peptide 36..125
FT mat_peptide 126..5093
FT /tag= b
FT /tag= c
FT /product= "Human mature 27875 ADAM-TS protein"
XX
XX WO200131034-A1.
XX
XX 03-MAY-2001.
XX
XX 25-OCT-2000; 2000WO-US29380.
XX
XX 25-OCT-1999; 99US-0426282.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Kapeller-Libermann R, White D;
XX
XX WPI; 2001-300513/31.
XX
XX P-PSDB; AA00913.
XX
XX Novel isolated polypeptide, 27875, a human ADAM-TS (a disintegrin and
XX metalloproteinase) useful for diagnosis and treatment of disorders of
XX bone, lung, heart, skeletal muscle such as osteoporosis, emphysema,
XX angina
XX
XX Claim 7; Page 115-123; 133pp; English.
XX
XX The present sequence is an alternative version of a cDNA encoding
XX 27875 protein, a human ADAM-TS (a disintegrin and metalloproteinase).
XX Metalloproteinase is a proteolytic enzyme involved in protein
XX maturation, protein degradation, tumour growth, metastasis and
XX angiogenesis. Nucleotides encoding 27875, 27875 protein and its
XX antibodies are useful for preventing, diagnosing and treating
XX 27875 metalloproteinase-related disorders. These disorders
XX include, disorders involving the lung such as congenital anomalies,
XX bronchial asthma, Goodpasture's syndrome, pulmonary alveolar proteinosis,
XX disorders involving heart such as heart failure, cardiac hypertrophy,
XX angina pectoris, myocardial infarction, chronic ischaemic heart disease,
XX disorders involving the skeletal muscle include tumours such as
XX rhabdomyosarcoma, disorders involving blood vessel such as hypertension,
XX atherosclerosis, vasculitis associated with other disorders, disorders
XX involving the testis and epididymis such as nonspecific epididymitis and
XX orchitis, gonorrhea, mumps, tuberculosis and syphilis, spermatocytic
XX seminoma, disorders involving the kidney such as membranoproliferative
XX glomerulonephritis, necrotising glomerulonephritis, renal artery
XX stenosis, chronic glomerulonephritis, sickle cell disease nephropathy,
XX urinary tract obstruction, disorders of the bone such as achondroplasia,
XX osteoporosis, Paget's disease, rickets, osteomalacia, osteonecrosis,
XX tuberculous osteomyelitis, osteoma, osteosarcoma, metastatic tumours,
XX and disorders involving thymus such as thymic cysts, thymic hypoplasia,
XX

```

CC thymomas, and Hodgkin's disease. 27875 sequence is used in gene therapy.  
 CC Note: This sequence SEQ.ID.NO.2 is stated as being similar to the  
 CC sequence shown in Fig 1 (See AAD04494). However these sequences differ  
 CC at several locations.

XX  
 SQ Sequence 5353 BP; 960 A; 1833 C; 1648 G; 911 T; 1 other;

# alignment\_scores:

Quality: 806.50 Length: 959  
 Ratio: 1.816 Gaps: 41  
 Percent Similarity: 46.298 Percent Identity: 28.050

## alignment block:

US-10-041-770-2 x AAD04475 ..

Align seg 1/1 to: RAD04475 from: 1 to: 5353

```

15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24
   : : : : : : : : : : : : : : : : : : :
631 GTGCTCCAAAGCACCCTGGAGTGCAAGTATCCAGAGCTGGAGTCTCGA 680
   : : : : : : : : : : : : : : : : : : :
24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
   : : : : : : : : : : : : : : : : : : :
681 CGGGACCGTGGGAGCAGCGGAGCAGTGGCGGCGCCACGCGTGAGGCG 730
   : : : : : : : : : : : : : : : : : : :
35 hrProThrGluGluGlyGlnGlyProGluGlyValTrpGlyProTrp... 50
   : : : : : : : : : : : : : : : : : : :
731 TCTACACCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780
   : : : : : : : : : : : : : : : : : : :
50 .....Gln 50
781 TAGCTATGCAAAATGGTGGAGTACCAACGAGCGGAGCGGAGTGGAGCG 830
   : : : : : : : : : : : : : : : : : : :
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
   : : : : : : : : : : : : : : : : : : :
831 TAGTGCTGACCATCATGAACATGGTGGTGGTGGTGGTGGTGGTGGTGG 880
   : : : : : : : : : : : : : : : : : : :
61 lyValGlyVal.....Gln 65
   : : : : : : : : : : : : : : : : : : :
881 CATGGGAACCCCATCCACATCACCATTGTGGCGCTGCTGCTGGGAAG 930
   : : : : : : : : : : : : : : : : : : :
66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80
   : : : : : : : : : : : : : : : : : : :
931 ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 980
   : : : : : : : : : : : : : : : : : : :
81 .....SerLeuProLeuP 85
   : : : : : : : : : : : : : : : : : : :
981 AGCTTCTGCAAGTGGCAGAAAGCATCAACATGAAGGGGATGCCCATCC 1030
   : : : : : : : : : : : : : : : : : : :
85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101
   : : : : : : : : : : : : : : : : : : :
1031 CCTGCACCATG..... 1041
102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118
   : : : : : : : : : : : : : : : : : : :
1042 ACACCTGCATCCTGCTGACCAAAA.....AGGACCTGTGT 1076
118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
   : : : : : : : : : : : : : : : : : : :
1077 GCAGCCATGAACCGGCGCTGTGAGACCTGGGACCTGCCATGTCCATGTG 1126
135 luGluThrGlnLuLeuArgAlaAlaArgSerArgLeuArgaspPro 151
   : : : : : : : : : : : : : : : : : : :
1127 CATGTGCCAGCGGACCGGAGCTGCAGCATCAACAGGAGACA..... 1167
152 IleLysProGlyMetPheGlyTrpGlyArgValProPheAlaLeuProLe 168
1167 ..... 1167
168 uHisArgAsnArgArgHisProArgSerProArgSerGluLeuSerL 185
   : : : : : : : : : : : : : : : : : : :
1168 ....CGGGCTGCGGCTTCACTGTAGCCACGAGCTCGGGGCACAG 1213

```

```

185 euIleSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201
   : : : : : : : : : : : : : : : : : : :
1214 TTTTGGCATTCAGCATGACGGAAGCGCAATGACTGTGAGCCCTGGGA 1263
   : : : : : : : : : : : : : : : : : : :
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
   : : : : : : : : : : : : : : : : : : :
1264 AACGACCTTTTCATCATGTCTCCACAGCTCCTGTACGAGCGCGCTCC 1313
   : : : : : : : : : : : : : : : : : : :
216 ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
   : : : : : : : : : : : : : : : : : : :
1314 ACCTGGTCCCGCTGCAGCGCCAGTATATACACAGGTTCCTTGTACCG 1363
   : : : : : : : : : : : : : : : : : : :
232 erProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
   : : : : : : : : : : : : : : : : : : :
1364 GTGGGCGCTGTGCTGGAGCACCCTCTCCCAAGGACATTTATCGACTCC 1413
249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258
   : : : : : : : : : : : : : : : : : : :
1414 CCTCGTGCCACCTGGCGCTCTCTATGATGTAAAGCCACAGTGGCGCTC 1463
258 rGlyThrGluProSerPro.....T 266
   : : : : : : : : : : : : : : : : : : :
1464 CAGTACGGGGGCTACTCTGCTTCTGCGAGGACATGGATATGTCTGCCA 1513
266 hrHisSerLeuGlyGlyGlyPheArgAlaSerProGlnProArg 282
   : : : : : : : : : : : : : : : : : : :
1514 CACACTCTGGTCTGTGGGA.....CCACCTGTC 1545
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg... 297
   : : : : : : : : : : : : : : : : : : :
1546 ACTCCAGCTGGATGCCCGCTGGAGCCACCCGCTGG...GGAGAAT 1592
298 .....ArgProAspProPheProSerValProArgG 308
   : : : : : : : : : : : : : : : : : : :
1593 AAGTGGTGTCTCAGTGGGAGTGGTACCCGTGGGCTTCCGGCCGAGG. 1641
308 lyArgGlyGlnGlyGlnGlyProTrp.....GlyThrGlyGly 321
   : : : : : : : : : : : : : : : : : : :
1642 .....CGTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1668
322 ThrProHisGlyProArgGluProAspProGlnHisProGlyAlaTr 338
   : : : : : : : : : : : : : : : : : : :
1669 GCGCCT...GGTCCA..... 1680
338 pleuProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355
   : : : : : : : : : : : : : : : : : : :
1691 .....TCTGTCCAGGAGCTGTGGCATGGCGGTACA 1711
355 heAlaProSer..Ser.....ProIleProArg..... 363
   : : : : : : : : : : : : : : : : : : :
1712 GAGCGCGAGCGGAGTGCACGCGACCTACCCCAATACAAAGGCAGAT 1761
364 ..CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCys 379
   : : : : : : : : : : : : : : : : : : :
1762 ACTGTGTGGTGAGCGAAGCGTTCGCTCTGCACCTGCAGCGCGCTGC 1811
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSe 396
   : : : : : : : : : : : : : : : : : : :
1812 CCTGCTGGCCCGCTCTCTCCGACAGTCCAGTCCAGCCACTTTGACGC 1861
396 rGlnGluPheMetGlyGlnLeuTyTrpGluProPheThrGluValG 413
   : : : : : : : : : : : : : : : : : : :
1862 TATGCTCTACAAAGGCGCAGCTGCACATGGTGGTGGCGTG.....GTCA 1905
413 lnGlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhe 429
   : : : : : : : : : : : : : : : : : : :
1906 ATGACGTGAACCCCTGCGAGCTGCACCTCGCGCGCGCGGAATGAGTACTTT 1955
430 TyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro... 445
   : : : : : : : : : : : : : : : : : : :
1956 GCCGAGAAGCTGGGGAGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2005

```

```

446 ...GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyC 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2006 CCGAGCCAGCGGAGCTCTGCATCAACGCATCTGTAGACAGCTGGGT 2055
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
461 yAspGlyLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2056 GTGACTTCAGATTGACTCCGCTGTATGAGGACCGCTGTGTGTGTGC 2105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
478 GlyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAsp 494
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2106 CAGGCAAGCGGTCCACCTGCCACACCGGTGAGCGGACCTTCGAGGAGC 2155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
494 gGlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAla 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2156 CGAGGCG...CTGGGTATGTGATGTGGGTGTATCCAGCGCGGCAC 2202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 euArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeu 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2203 GCGAGTCCGCATCCAGAGGTTCGAGGCTGCCACATCTCTGGCAGTG 2252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
528 ArgGlyProGly...GlyArgSerIleLeuAsnGlyAsnTrpAlaValas 543
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2253 CGGAGTGAGGACCGGAGAGTACTTCCCAATGGTGGTGAGCATCCA 2302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 pProGlySerTyrArgAlaGlyGlyThrValPheArgTyrAsnArgP 560
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2303 GTGGAACGGGGACTCCAGGTGGCAGGACACCTTCATCATCGCACGC 2351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
560 roProArgGluGluGlyLysGlyGlySerAlaGluGlyProThr 576
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2352 .....AGGGCAACTGGAGAACCTCAGCTCCCGGGTCCACCC 2390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
577 ThrGlnProValAspValTyrMetIlePheGlnGluGluAsnProGlyVa 593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2391 AAGGAGCCTGTCTGGATCCAGCTGTGTTCAGAGAGACAACCTGGGT 2440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
593 lPheTyrGlnTyrValIleSerSerProProIleLeuGluAsnProT 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2441 GCATCAGGAGTAC..... 2453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
610 hrProGluProProValProGlnLeuGlnProGluLeuArgValGlu 626
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2453 ..... 2453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeu 643
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2454 .....ACCATCCA 2461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
643 nArgGlnValArgIleProGlnMetProAlaProHisProArgThrP 660
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2462 CAGGGAG.....GCAGGTGGCCACGACGAGGTCC 2490
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 roLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCys 676
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2491 CG.....CGGCCGTCTCTCTGCATATGGGCCCTGGACCAAGTGC 2534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
677 SerAlaSerCysGlyLysValTrpArgProIlePheLeuCysIleSe 693
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2535 ACAGTCACCTGGCGCAGAGGTGTGCAGAGGCAGAAATGTCTACTGCTGGA 2584
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAla 710
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2585 GCGGCAGGAGGG...CCCGTGGACGAGGACACGTGTACCCCTGGGGC 2631
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
710 rgProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2632 GGCCTGATGACCAACAGAGAGTGCAGGAGCAGGCTGCTGCCAGG 2681
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...Gl 742
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2682 TGTGGCAGAGGTGAGTGGCAGTGTCTCAGCTCTCGGGCGGGGG 2731
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
742 yThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGly 758

```

```

2732 CCTTCCCGCGGCGCTCTGCATCCGCGGTGGGTGGATGAGC 2781
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
758 lySerSerValProGluArgCysGlyHisLeuProArgProAsnIle 774
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2782 AGAGCGCCTGGAGCCACCCCGCTGTGAACACCTCCCGCGGCCCTACT 2831
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValGl 789
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2832 GAAACCCCTTCACCGCATGTACCTGTCCGCGCCACCTGGGTGTGG 2881
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerA 806
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2882 GAAC...TGGTCTCAGTGTGACATGTGGGAGGGGCACCTCAGCGCC 2928
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2929 GAAATGTCTCTGCACCAATGACACCGGTGTCCCTGTGAGGAGGCCAG 2978
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
823 CysAlaSerGlyProProGlnProProSerArgGluAlaCysaspMetGl 839
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2979 .....CAGCCAGCCAGCAGTACCTGCTCTCTGCC 3010
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
839 yProCysThr...ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3011 ACTCTGTCTGGTGGCCCTGGGCACAC.....TGGGCCCTGAAGGCTCAG 3054
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
855 SerProGluProProAlaIleSer 862
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3055 GCAGGGGTCTCTCCAGCCACGAGC 3078
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: /SIDS1/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH49372
seq_documentation_block:
ID AAH49372 standard; DNA; 5338 BP.
XX AAH49372;
XX DT 07-DEC-2001 (first entry)
XX DE Human metalloprotease MPTS-19 encoding DNA.
XX KW Metalloprotease; MPTS-10; human; thrombospondin domain; spondylitis;
XX KW aggrecan associated disease rheumatoid arthritis; osteoarthritis;
XX KW osteopathic; antiarthritic; antipsoriatic; antirheumatic; psoriasis;
XX KW sport injury; joint trauma; fibrosis; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..5073
XX FT /*tag= a
XX FT /product= "MPTS-19"
XX PN DE10107360-Al.
XX PD 06-SEP-2001.
XX PF 16-FEB-2001; 2001DE-1007360.
XX PR 18-FEB-2000; 2000US-184152P.
XX PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX PI Heller RA, Klonowski P, Zuo F;
XX DR WPI; 2001-607955/70.
XX DR P-PSDB; AAB86949.
XX PT Novel metalloprotease containing a thrombospondin domain (MPTS protein)
XX PT is useful to treat aggrecan associated disease including rheumatoid
XX PT arthritis and osteoarthritis
XX

```



PS Claim 1b; Page 39-43; 66pp; German.  
 XX This invention describes a novel metalloprotease containing a  
 CC thrombospondin domain (MPS protein) which is useful to treat aggrecan  
 CC associated disease including rheumatoid arthritis and osteoarthritis.  
 CC The products of the invention have osteopathic, antiarthritic,  
 CC antiproliferative and antirheumatic activity. MPS may be used to identify a  
 CC modulator of its activity, e.g., an agonist or antagonist. Such  
 CC compounds, as well as MPS itself may be used to treat disease  
 CC associated with MPS activity or aggrecan degradation, particularly  
 CC osteoarthritis, rheumatoid arthritis, psoriasis, spondylitis, sport  
 CC injury, joint trauma or fibrosis. This sequence encodes the human  
 CC MPS-19 metalloprotease described in the method of the invention.  
 XX  
 SQ Sequence 5338 BP; 974 A; 1825 C; 1638 G; 901 T; 0 other;

alignment\_scores:  
 Quality: 804.50 Length: 959  
 Ratio: 1.808 Gaps: 41  
 Percent Similarity: 46.403 Percent Identity: 27.946

alignment\_block:

US-10-041-770-2 x AAH49372 ..

Align seg 1/1 to: AAH49372 from: 1 to: 5338

```

15 LeuSerLeuProGlnLeuCys.....LeuAs 24
   : : : : : : : : : : : : : : : : : :
608 GTGCTCCAAAGACCTGTGGAGTGCAAGTATACCCAGAGCTGGAGCTCGA 657
   : : : : : : : : : : : : : : : : : :
24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
   : : : : : : : : : : : : : : : : : :
658 CGGAGCGTGGGAGCAGCGGAGAGTGGCGCGCCAGCGCTGAGCG 707
   : : : : : : : : : : : : : : : : : :
35 hrProThrGluGlyGlnGlyProGluGlyValTrpGlyProTrp... 50
   : : : : : : : : : : : : : : : : : :
708 TCTACACCGAGCGTGGTCAAGAGAGAGTGGTGGAGACCTGGTGG 757
   : : : : : : : : : : : : : : : : : :
50 ..... 50
758 TAGCTATGCCAAATGGTGGAGTACCAGGAGACGCCGAGTTGAGAGC 807
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
   : : : : : : : : : : : : : : : : : :
808 TATGTGCTGACCATCATGAACATGTGTGGCTGGCTGTTTCATGACCCAG 857
   : : : : : : : : : : : : : : : : : :
61 lyValGlyVal.....Gln 65
   : : : : : : : : : : : : : : : : : :
858 CANTGGGAACCCATCCACATCACCATTGTGCGCTGGTCTGCTGGAG 907
66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80
   : : : : : : : : : : : : : : : : : :
908 ATGAGGAGGAGGACCTAAAGATCAGCCACCATGCAGACAACACCCGGAAG 957
81 .....SerLeuProLeuP 85
958 AGCTTCTGCAATGGCAGAAAGCATCAACATGAAGGGGATGCCCATCC 1007
85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101
1008 CTGCACCATG..... 1018
102 ProArgProGlnThrSerProGluThrLeuProLeuTrpArgThrGlnSe 118
1019 ACACGTGCCATCTGCTGCCAGAAA.....AGGACCTGTGT 1053
118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
   : : : : : : : : : : : : : : : : : :
1054 GCAACCATGAACCGGCGCTGTGACCCCTGGGACCTGTCCTCCATGTGGCGG 1103
135 luGluThrGlnGluLeuArgAlaAlaArgSerArgLeuArgAspPro 151
   : : : : : : : : : : : : : : : : : :

```

```

1104 CATGTGCCAGCGCAGCGAGCTGCAGCATCAACAGGAGCA..... 1144
152 IleLysProGlyMetPheGlyTyrglyValProPheAlaLeuProLe 168
1144 ..... 1144
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
   : : : : : : : : : : : : : : : : : :
1145 ...CGGCGCTGGCGGCTTACATGTAGCCACGAGCTCGGCACAG 1190
185 euLeuSerSerArgGlyGluAlaLeuProSerProThrProArgAla 201
   : : : : : : : : : : : : : : : : : :
1191 TTTTGGCATTCAGCATGACGAGGCGCAATGACTGTGAGCCGTTGGGA 1240
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
   : : : : : : : : : : : : : : : : : :
1241 AACGACCTTTTCATCATGCTCCACAGCTCTGTACGACGCCGCTCCCTC 1290
216 .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
   : : : : : : : : : : : : : : : : : :
1291 ACCTGGTCCCGCTGCAGCGCCATATACACAGGTTCTTGTACCGTGG 1340
232 erProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
   : : : : : : : : : : : : : : : : : :
1341 GTGGGCGCTGTGCTGGAGCAGCCCTCTGCCAAGGACATTATCGACTTC 1390
249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258
   : : : : : : : : : : : : : : : : : :
1391 CCTCGTCCACCTGGCGTCTCTATGATGAAGCCACGAGTGGCGCTC 1440
258 rGlyThrGluProProSerPro.....T 266
   : : : : : : : : : : : : : : : : : :
1441 CAGTACGGGCGCTACTCTGCTTCTGCGAGGACATGATAATGTCGCCA 1490
266 hrHisSerLeuGlyGluGlyPheArgAlaSerProGlnProArg 282
   : : : : : : : : : : : : : : : : : :
1491 CACACTCTGGTCTCTGTGGGA.....CCACCTGTCT 1522
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg.. 297
   : : : : : : : : : : : : : : : : : :
1523 ACTCCAAGCTGGATGCAGCGCTGAGCGCACCGCGTGTG...GGAGAAT 1569
298 .....ArgProAspProPheProSerValProArg 308
   : : : : : : : : : : : : : : : : : :
1570 AAGTGGTGTCTCAGTGGGAGTGCCTACCCGTGGCTTCCGCGCCGAGG. 1618
308 lyArgGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321
   : : : : : : : : : : : : : : : : : :
1619 .....CGGTGGATGTGGCTGTGCTGCTGCTGGA 1645
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338
   : : : : : : : : : : : : : : : : : :
1646 GCGCT...GGTCCA..... 1657
338 pLeuProLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeup 355
   : : : : : : : : : : : : : : : : : :
1658 .....TCTGTCCAGGAGCTGTGCATGGCGCTACA 1698
355 heAlaProSer...Ser.....ProIleProArg..... 363
   : : : : : : : : : : : : : : : : : :
1689 GAGCGCGAGCGGAGTGCAGCCAGCTACGCCCAATAACAAAGGACAGAT 1738
364 ..CysSerGlyCysSerGluGlnLeuArgAlaCysSerGlnAlaProCys 379
   : : : : : : : : : : : : : : : : : :
1739 ACTGTGGGTGAGCGGAAGCGCTTCCGCTCTGCAACCTGCAGCGCTGC 1788
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAnSe 396
   : : : : : : : : : : : : : : : : : :
1789 CTGTGTGGCGCCCTCTCTCCGCCAGCTCCAGTGCAGCCACTTTGACGC 1838
396 rGlnGluPheMetGlyGlnLeuTrpGlnTrpGluProPheThrGluValG 413
   : : : : : : : : : : : : : : : : : :
1839 CATGCTCTACAAGGCGGCTGCACACATGGGTGCCCGTG.....GTCA 1882

```





```

354 upheAlaProSer.SerProIleProArg..... 363
|||||
1706 CTTCTAGCGCTCACTCGACACCCAGCCCAACCATCGGGGCAAGTAC 1755
364 CysSerGlyLeuSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380
|||||
1756 TGCTGGGTGAGAAAGCGGACCGCTCCTCAACACGATGACTGTCC 1805
380 oProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397
|||||
1806 CCTTGGCTCCAGGACTTCAGAAAGTGCAGTGTCTGAATTGACAGCA 1855
397 InGluPheMetGlyGlnLeuTyrlntPrGluProPheThrGluValGln 413
|||||
1856 TCCTTTTCGTGGGAATCTCAAGTGGAAACGTAC...CGGGGAGGG 1902
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTy 430
|||||
1903 GCGGTGAGCCCTCGCTCAGTCCAGTCCCTAGCGGAAGGCTTCAACTTCA 1952
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447
|||||
1953 CACGGAGAGGGCGGACCGTGTGGACGGACACCTGCCGTCCAGACA 2002
447 laProaspIleCysValAlaIArgCysLeuSerProGlyCysaspGly 463
|||||
2003 CGGTGACATTCCTCAGTGGCGAATGAAGCACGTGGGCTGCGACCGA 2052
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAs 480
|||||
2053 GTCTGGCTCCGACCTCGCGGAGGACAAAGTCCGAGTGTGGCGGTGA 2102
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497
|||||
2103 CGGCAGTGTGGACCACTCAGGCGCTTTCACCCAGCCTCACCTG 2152
497 roLeuGlyTyrlntLysIleLeuTyrlntProAlaGlyAlaLeuArgLeu 513
|||||
2153 GGGCGGTACGAGGATGCTGTGATTCCCAAGGCTCCGTCACATC 2202
514 GlnIleAlaGlnArgProSerSerAsnTyrlntLeuAlaLeuArgGlyPr 530
|||||
2203 TTCATCCAGATCTGAACCTCTCTCAGTCACTTGGCCCTGAAGGAGA 2252
530 oGlyGlyArgSerIleAsnGlyAsnTrpAlaValAspProGlyLys 547
|||||
2253 CCAGGAGTCTCTGCTGGAGGGGTGCGCGGACCCCAACCCCAACC 2302
547 erTyrlArgAlaGlyThrValPheArgTyrlntAsnArgProProArgGlu 563
|||||
2303 GTCTGCTCTAGTGGACCACTTCACTGCGACAGGGGCGGACGAC 2352
564 GluGlyLysGlyGluSerLeuSerAlaGluGlyProThrThrGlnProVa 580
|||||
2353 GTC.....CAGACCTCGAAGCCTGGACCGGATTAATGATCTCT 2393
580 lAspValTyrlMetIlePheGlnGluGlnAsnProGlyValPheTyrlnt 597
|||||
2394 CATGCTCATGCTGTGCGCGGACCGAGTGCCTGCCCTCCGCTACCGCT 2443
597 yrValIleSerSer.ProProIleLeuGluAsnProThrProGluPr 613
|||||
2444 TC.....AATGCCCATCGCGGTGACTCGTCGCC....CTACTCC 2484
613 o.....ProValProGlnLeuGlnProGlu..... 621
|||||
2485 TGGCATTATGCGCCCTGGACCAAGTGTGCGGCCACGATGTCAGGCGGTAG 2534
622 ..... 626
2535 CCAGGTGACGGCGGTGAGTGCCTCCCAACCAAGCTGGACAGCTCCGCGGTG 2584

```

626 upProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
:||||| :|  
2585 CCCCCACTACTGAGTCCACAGCAAGCT.....GCC 2619  
643 InArgGlnValArgIleProGln.MetProAlaProHisProArgTh 659  
:||||| :|  
2620 AAAAGGAGCGCGCTGCACACGAGCCTTGCCTCCAGAC..... 2661  
659 rProLeuGlySerProAlaAlaTyrlntLysArgValGlyHisSerAlaC 676  
:||||| :|  
2662 .....TGGGTTGTAGGAAGTGGTCTGCTCT 2686  
676 ySerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysile 692  
:||||| :|  
2687 GCAGCGCAGCTGCATCGAGCGTGGCAGCCCTCGGTGCTGTGCAG 2736  
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706  
:||||| :|  
2737 CGCGCGTCTCTCGCGGAGGAGGAGCGCTGGACGACAGCGATGCC 2786  
706 aAlaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProC 723  
:||||| :|  
2787 GCAG...CGCGCCACCTGTACTG...GAGGCTGCCACGCCCACTT 2830  
723 ysProProTyrlntGluAlaGlyGluTrpThrSerCysSerArgSerCys 739  
:||||| :|  
2831 GCCCTCCGAGTGGCGCCCTCGACTGCTGTGAGTCACCCCGAGCTGC 2880  
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyG 756  
:||||| :|  
2881 GGGCGGCGCTCCGCCACCGGTGCTTTCAGAGAGC.....GCAGA 2924  
756 yGlyGlySerSerValProGluArgCysGlyHisLeuProArgProA 773  
:||||| :|  
2925 CCACCGCGCGAGTGCCTCGCGGCACTGTACCCCGCGCCCAAGCCAC 2974  
773 snIleThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787  
:||||| :|  
2975 CGCGCACCATGCTGCAACTGCCCGCTGCCCCCGCGCGCTGGGTG 3024  
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804  
:||||| :|  
3025 GCTGCGGAG...TGGGTGAGTGTCTGCACAGTGGCGGTGGGSCACG 3071  
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821  
:||||| :|  
3072 GCAGGCTCGTGCCTGCACGACGACACGCGGCGCAG.....GCGTCG 3115  
821 InGluCysAlaSerGlyProGlnProProSerArgGluAlaCysasp 837  
:||||| :|  
3116 ACAGTGCACGAGGCG...CTGCGCGCGCCACACGACGAGTGTGAG 3162  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAF82157  
seq\_documentation\_block:  
ID AAF82157 standard; cDNA; 3312 BP.  
XX AAF82157;  
AC AAF82157;  
XX 02-JUL-2001 (first entry)  
DT 02-JUL-2001 (first entry)  
XX Human ADAM type metal protease MDT2 encoding cDNA SEQ ID NO:11.  
DE Human; a disintegrin and metalloprotease type metal protease; MDT2;  
KW MDT2; MDT23; ADAM type metal protease; cytosolic; antiarthritic;  
KW cancer; arthritis; arthrosis deformans; ss.  
XX Homo sapiens.  
OS  
PH Key  
FT 1..3312  
FT CDS  
FT /tag= a  
FT /product= "MDT2"



Wed Jul 24 11:39:30 2002

364 CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380  
1756 TGTCTGGGTGAGAAAGCGCGACCGTCTCTGCAACACAGGATGACTGTCC 1805  
380 OProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397  
1806 CCTGGCTCCCGAGCATTCAGAGAAGTGCACTGTCTGAATTTGACAGCA 1855  
397 InGluPheMetGlyClnLeuTyrrGlnTrpGluProPheThrGluValGln 413  
1856 TCCCTTTCCGTGGGAATTCACAAGTGGAAACGTAC...CGGGAGGG 1902  
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTy 430  
1903 GCGCTGAGGCGCTGCTCGTCACTGCTACGGGAAGSGCTTCACTTCTA 1952  
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447  
1953 CACGAGAGGCGCGACGCGTGGTGGACGGGACACCTCGCGTCCAGACA 2002  
447 IapCAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463  
2003 CGGTGGACATTTGCTAGTGGCAATGCCAAGCAGCGTGGCTGCGACCGA 2052  
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAs 480  
2053 GTCTCTGGGTCCGACTCGCGGAGACAAAGTGCAGATGTGTGGCGGTGA 2102  
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497  
2103 CGCGAGTGCCTCGAGACATTCGAGGGCGTCTTCAGCCACGCTCACCTG 2152  
497 roLeuGlyTyrrGlnLysIleLeuThrIleProAlaGlyAlaLeuArgLeu 513  
2153 GGGCGGGTACGAGAGATGCTGTGTGATTTCCAAAGSGCTCGTCCACATC 2202  
514 GlnIleAlaGlnLeuArgProSerSerAsnTyrrLeuAlaLeuArgGlyPr 530  
2203 TTCATCCAGANTCTGAACCTCTCTCTAGTCACATGCCCCCTGAAGGAGA 2252  
530 oGlyGlyArgSerIleAsnGlyAsnTrpAlaValasProProGlyS 547  
2253 CCAGGAGTCCCTGCTCTGGAGGGGCTCGCCGGAGACCCCGAGCCCGACC 2302  
547 erTyrrArgAlaGlyGlyThrValPheArgTyrrAsnArgProProArgGlu 563  
2303 GTCTGCCTCTAGCTGGGACCACTTCACTGCGACAGGGCCAGACAGC 2352  
564 GluGlyLysGlyGluSerLeuSerAlaGluGlyProThrThrGlnProVa 580  
2353 GTC.....CAGAGCTCGAAGCCCTCGGACCGGATTAATGATCTCT 2393  
580 IaspValTyrrMetIlePheGlnGluGluAsnProGlyValPheTyrrGlnT 597  
2394 CATCGTCATGGTGTGGCCGACCGAGCTGCCCTGCCCTCGCTACCGCT 2443  
597 yrValIleSerSerProProIleIleLeuGlnAsnProThrProGluPr 613  
2444 TC.....AATGCCCATCGCCCTGACTCGCTGCCCC...CCTACTCC 2484  
613 o.....ProValProGlnLeuGlnProGlu..... 621  
2485 TGGCACTATGCGCCCTGGACCAAGTGTGGCCCACTGTGACGGCGGTAG 2534  
622 .....Ile.LeuAaGvalGl 626  
2535 CCAGTGCAGGGGTGGACTGCGCAACACGCTGGACAGCTCCGGGTGCG 2584  
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2585 CCGCCCACTACTGCACTGCCCAACGACGCT.....GCC 2619

```

643 lnArglnValargIleProGln_MetProAlaProProHisProArgTh 659
||||| :|||::: ::| |::| |||||
2620 AAAAGCAGCGCGCTGCAACACGAGCGTTCGCCCTCCAGAC..... 2661
||||| :|||::: ::| |::| |||||

659 rProLeucylSerProAlaAlaTyTrpLysArgValGlyHisSerAlac 676
||||| :|||::: ::| |::| |||||
2662 .....TGGGTTGTAGGAACTGGTGCCTCT 2686
||||| :|||::: ::| |::| |||||

676 ysSerAlaserCysGlyLysGlyValTrpArgproIlePheLeuCysIle 692
||||| :|||::: ::| |::| |||||
2687 GCAGCGCAGCTGCATGCAGCGCTGCAGCGCTCGCTCGCTGTCGCAG 2736
||||| :|||::: ::| |::| |||||

693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706
||||| :|||::: ::| |::| |||||
2737 GCCCGCGTCTGCGCGGAGAGAGCGCTGAGCAGCGCATGCC 2786
||||| :|||::: ::| |::| |||||

706 aaLaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProc 723
||||| :|||::: ::| |::| |||||
2787 GCAG...CGCGCCCCACTGTACTG...GAGGCTGCCACGCGCCCACATT 2830
||||| :|||::: ::| |::| |||||

723 ysProProTyrrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739
||||| :|||::: ::| |::| |||||
2831 GCCCTCCGAGTGGCGGCCCTGCGACTGCTGAGTGCACCCCGCAGCTGC 2880
||||| :|||::: ::| |::| |||||

740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGl 756
||||| :|||::: ::| |::| |||||
2881 GGGCGGGGCTCGGCACCGCGCTGGTCTCTTGCAGAGC.....GCAGA 2924
||||| :|||::: ::| |::| |||||

756 yGlyGlySerSerValProProGluArgCysGlyHisLeuProoArgProA 773
||||| :|||::: ::| |::| |||||
2925 CCACGCGCGCAGCTGCCCGCGCGCTGCTCACCGCGCGCAAGCCAC 2974
||||| :|||::: ::| |::| |||||

773 snileThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787
||||| :|||::: ::| |::| |||||
2975 CGGCCACCATTGCTGCAACTTGCAGCTGCGCGCTGCCCGCGCGCTGGTG 3024
||||| :|||::: ::| |::| |||||

788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804
||||| :|||::: ::| |::| |||||
3025 GTGTGGCAG...TGGGGTAGTGTCTGCACATGCGCGCTCGCGGAGCG 3071
||||| :|||::: ::| |::| |||||

804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821
||||| :|||::: ::| |::| |||||
3072 GCAGCGCTCGGTGCGTGCACCGACACACAGGGCCAG....GCGTGC 3115
||||| :|||::: ::| |::| |||||

821 lnGluCysAlaserGlyProProGlnProProSerArgGluAlaCysAsp 837
||||| :|||::: ::| |::| |||||
3116 ACAGATGTCAGGAGGCC...CTCGCGCGCGCCACACAGCAGCTGTGAG 3162
||||| :|||::: ::| |::| |||||

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:A
seq_documentation_block:
ID AAH43769 standard; DNA; 3776 bp.
XX
AC AAH43769;
XX
DT 30-JAN-2002 (first entry)
XX
DE ADAMTS-E DNA.
```

seq name: /SIDSL/qcadata/hold-geneseq/geneseq-emb1/NA2002.DAT:AAH43769

seq\_documentation\_block:

ID AAH43769 standard; DNA; 3776 BP.

XX

AC AAH43769;

XX

DT 30-JAN-2002 (first entry)

XX

DE ADAMTS-E DNA.

ADAMTS-E; disintegrin; metalloprotease; thrombospondin domain; stroke;  
 ADAMTS-E3; splice variant; osteoarthritis; rheumatoid arthritis; pain;  
 inflammatory bowel disease; Crohn's disease; emphysema; infertility;  
 acute respiratory distress syndrome; spinal cord injury; head trauma;  
 chronic obstructive pulmonary disease; Alzheimer's disease; asthma;  
 organ transplant toxicity; rejection; cachexia; allergy; cancer;  
 tissue ulceration; restenosis; periodontal disease; Parkinson's disease;  
 epidermolysis bullosa; osteoporosis; atherosclerosis; aortic aneurysm;  
 congestive heart failure; myocardial infarction; cerebral ischaemia;  
 neurodegenerative disease; Huntington's disease; peripheral neuropathy;  
 cerebral amyloid angiopathy; amyotrophic lateral sclerosis; migraine;  
 multiple sclerosis; ocular angiogenesis; corneal injury; depression;  
 macular degeneration; abnormal wound healing; burns; diabetic shock; d

OS Homo sapiens.

XX Key Location/Qualifiers

XX FH 139..3453

XX CDS /\*tag= a

XX FT /product= "ADAMTS-E"

XX EP1149903-A1.

XX PN

XX PD 31-OCT-2001.

XX XX

XX 26-APR-2001; 2001EP-0303818.

XX PF

XX PR 26-APR-2000; 2000US-199924P.

XX PP (PFZ ) PFIZER PROD INC.

XX PI Buckblinder L, Mitchell PG, Walsh RT, Wachtmann TS;

XX DR

XX WIPI; 2002-019321/03.

XX PS P-PSDB; AAB47719.

XX PT Novel polynucleotide encoding ADAMTS-E, a disintegrin and metalloproteinase containing thrombospondin domain, and ADAMTS-E2 polypeptide useful for treating cerebral ischemia, osteoporosis, cachexia, allergies and asthma .

XX PT Claim 1; Page 16-17; 48pp; English.

XX CC This sequence encodes ADAMTS-E which is a disintegrin and metalloprotease containing a thrombospondin domain. ADAMTS-E2 is a splice variant of ADAMTS-E. The ADAMTS-E DNA is useful for diagnosing a disease or susceptibility to a disease in a subject relative to expression or activity of ADAMTS-E in a subject. ADAMTS-E and the splice variant may be used in the treatment of osteoarthritis and rheumatoid arthritis, inflammatory bowel disease, Crohn's disease, emphysema, acute respiratory distress syndrome, asthma, chronic obstructive pulmonary disease, Alzheimer's disease, organ transplant toxicity and rejection, cachexia, allergy, cancer, tissue ulcerations, stenosis, periodontal disease, epidermolysis bullosa, osteoporosis, loosening of artificial joint implants, atherosclerosis, aortic aneurysm, congestive heart failure, myocardial infarction, stroke, cerebral ischemia, head trauma, spinal cord injury, neurodegenerative diseases, autoimmune depression, Huntington's disease, Parkinson's disease, migraine, depression, peripheral neuropathy, pain, cerebral amyloid angiopathy, nootropic, or cognition enhancement, amyotrophic lateral sclerosis, multiple sclerosis, ocular angiogenesis, corneal injury, macular degeneration, abnormal wound healing, burns, infertility or diabetic shock.

XX SQ Sequence 3776 BP; 710 A; 1253 C; 1202 G; 611 T; 0 other;

alignment\_scores:

	Quality:	Length:
	796.00	954
Ratio:	1.697	Gaps: 44
Percent Similarity:	49.161	Percent Identity: 28.407

alignment\_block:

US-10-041-770-2 x AAH43769 ..

Align seg 1/1 to: AAH43769 from: 1 to: 3776

4 TrpThrGlyArgProTyrLeuLeuLeuLeuSerLeuProGl 20  
|||||  
742 TGGAAGGGCGGCCATGG...TGCGTCGGACCTTGAAGCCACCGCTGC 788  
|||  
20 nLeuCysLeuaspGlnGluValLeuSerGlyHisSerLeuGlnThrProt 37  
||| |||  
789 CAGGCCCTGGGGAATGAA.....GlyProGlu.GlyVa 46  
||| |||  
37 hrGluGluGlyGln.....

```

alignment_scores:      Length: 954
                        Quality: 796.00
                        Ratio: 1.697
                        Gaps: 44
                        Percent Identity: 28.407

alignment_block:
US-10-041-770-2 x AAH43769 ..

Align seg 1/1 to: AAH43769 from: 1 to: 3776

4 TrpThrGlyArgProTlpLeuTyrrLeuLeuLeuLeuSerLeuProG1 20
|||||
742 TGGAAAGGGCGGCCATGG...TGCTCGGACCTTGAAGCCACCGCGCTGC 788
|||||

20 nLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuGlnThrProt 37
|||||
789 CAGGCCCTGGGGAATGAA.....A 808
|||||

37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46
|||||

```





seq\_documentation\_block:

seq\_documentation\_block:  
ID AAF63445 standard: cDNA: 3885 bp.

XX  
AC AAF63445:

XX  
XX  
XX

14-MAY-2001 (first entry)

XX  
XX

Human ADAMTS-10 cDNA sequence.

XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;  
KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIC; angiogenesis;  
KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;  
KW Metastasis; embryogenesis; egg implantation; ADAMTS-10; ss.

KW metastasis; embryog  
YY

XX Homo sapiens.

XX  
XX  
XX

PN WO200111074-

XX  
XX

PD 15-FEB-2001.

XX  
XX

PF 03-AUG-2000; 2000

XX

PR 06-AUG-1999; 99US-

[illegible]PA (CLEV-) CLEVELAND CL  
(ADTF/) ADTF S C  
PA (ADTF/) ADTF S C

PA (APTE//) APTE S S:  
PA (HITS//) HITSKATNE

PA (HUKS/) HURSKAINEN I L.  
PA (HIRO/) HIROHATA S.  
XX  
PI Apte SS, Hurskainen TL, Hirohata S;

XX  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  

WPI: 2001-159978/16.  
P-PSDB; AAB72288.

Murine and human 'A Disintegrin-like And Metalloprotease domain with Thrombospondin type I motifs' proteins and the nucleic acids encoding them, useful for treating e.g. tumours, inflammation and arthritis -

Claim 15; Fig 9; 181pp; English.

This invention relates to murine and human ADAMTS-N (A disintegrin-like and metalloprotease domain with thrombospondin type I motifs) proteins, designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the invention are cDNA sequences encoding the proteins, and antibodies specific for the proteins. The nucleic acid sequences and proteins may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate ADAMTS-N expression. Disorders that may be treated using the nucleic acids, proteins and antibodies include, for example, tumour cachexia, inflammation, dematoparaxia in cattle or Ehlers-Danlos syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage in arthritic (both inflammatory and non-inflammatory) disease, angiogenesis, tumour growth and metastases, and they may also be used for controlling embryogenesis and implantation of fertilised eggs. The present sequence represents cDNA encoding human ADAMTS-10.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Sequence 3885 BP; 795 A; 1209 C; 1172 G; 707 T; 2 other;  
SQ

alignment scores:

Quality:	790.00	Length:	1013
----------	--------	---------	------

Ratio:	1.649	Gaps:	49
--------	-------	-------	----

Percent Similarity:	47.285	Percent Identity:	27.739
---------------------	--------	-------------------	--------

alignment\_block:

US-10-041-770-2 x AAF63445

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

.....

4 TrpThrGlyArgProTrpLeuTyrLeuLeuLeuSerLeuProG1 20

20 nLeuCysLeuaspGlnGluValLeuSerGlyHisSerLeuGlnThrProT 37  
 ||||| :|||  
 585 CAGACCCCTGGGAATGAA.....A 604

37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46  
 |||||:|||||  
 605 CAGAGCGTGGCCAGCCAGGCTGAACGGATCGGTGAGCGAGAGCGCTAC 654  
 :||| :|||  
 46 lTTPGlyProTrpVal..... 51  
 :||| :|||  
 655 GTGAGACCATGGATGGCTGTGACAAAGATGATGGTGGCTATCAGGGCG 704

52 .....GlnTrpAlaSerCysSerGlnProCysGlyVal.GlyValGlnArg 66  
 :|||:|||||  
 705 CCGGATGTGGACGATATGCTCTGCCCATCATGACATGTTGGCAAC 754  
 :||| :|||  
 67 ArgSerArgThr..... 70  
 :|||:|||||  
 755 TTTTCCAGGATCGAGTCTGGGAAGCACCGTTAACTCTCGTAACTCGC 804

71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85  
 :|||:|||||  
 805 CTCATCTGTCTCAGGAGGACCGCCACTCTGGAGATCA..... 844

85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
 :||| :|||  
 845 .CCACCATCGCGGGAAGTCCCTAGACAGCTCTCTAAGTGGCAGAAAT 892  
 :||| :|||  
 102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThr..... 116  
 :||| :|||  
 893 CCATCGTGAACACAGACGGGCATGGCAATGCCATTCAGAGAAGCGGTG 942

116 ..... 116  
 :||| :|||  
 943 GCTAACCATGACACAGAGTGCTCATCAGACGGTATGACATCTGCATCTA 992

117 .....GlnSer.ArgGlyArgGlyGlyProLeuArgGlyProAl 129  
 :||| :|||  
 993 CAAGAACAACCCCTGCGGCACACTAGGCGCTGGCCGCTGGCGGAATGTG 1042

129 aserHisLeuGlyArgGluGlnThr.....GlnGluLeuArgA 142  
 :||| :|||  
 1043 TGAG.....CGCAGAGAGAGCTCCACCGTCATGAGACATCTGGCT 1083  
 :||| :|||

142 laAlaArgSerArgLeuArgAspProLeysProGlyMetPheGly 158  
 :||| :|||  
 1084 GCCAAGCGTTCA..... 1097

159 TyrGlyArgValProPheAlaLeuProLeuHisArgAsnArgArgHisPr 175  
 :||| :|||  
 1098 .....CCATTCCACGAGATCGGCACACAT 1123

175 oArgSerProProArgSerGluLeuSerLeuHisSerArgGlyGluG 192  
 :||| :|||  
 1124 TCGCATGAACCA.....TGACGCGCTGGGAAC 1152

192 luAlaLeProSerProThrProArgAlaGluProPheSerAlaAsnGly 206  
 :||| :|||  
 1153 AGCTGTGG.....GGCCCGTGGTCAAGACCCAGCCAAAGCTCATGGC 1193

209 SerProGln...ThrGluLeuProProThrGluLeuSerValHisThrPr 224  
 :||| :|||  
 1194 TGCCCATCATTAACATGAAGACCAACCCATTCTGGTGGTCTCATCTGCAACC 1243

224 o.....SerProGlnAlaGluProLeuSerProGluThrAlaGlnThrG 239  
 :||| :|||  
 1244 GTGACTACATCACCAGCTTTCTAGACTCGGCGCTCTGCCTGAAC 1293

239 luValAlaProArgThrArgProAlaProLeuArgHisHisProArgAla 255  
 :||| :|||  
 1294 AACGGGCCCCAGACAGGACTTTG.....GTA 1322

256 GlnAlaSerGlyThrGluPro.....ProSerProTh 266

1323 CCCGACAGTGCACCGGGCCCAAGCCTACGATGCAGATGAGCAATGCCGCT 1372  
266 rHisSerLeuGlyGlu.....GlyGlyPhePheA 276  
1373 TTCAGCATGGAGTCAAAATCGCTCAGTGTAAATACAGGGGAGTCTGCAGC 1422  
276 rGAla.....SerProGlnPro.....ArgArgPro 284  
1423 GAGCTGTGGTGTCTGAGCAAGAGAACCGGTGCATCACCAACAGCATCCC 1472  
285 SerSerGlnGlyTTPAlaSerProGlnValAlaGlyArgArgPro..... 299  
1473 GCGCCCGAGGCGAGCTGTCCAGAGCCACACCATCGACAAAGGGTGGT 1522  
300 .....AspProPheProSerValProArgGlyGly 311  
1523 GCTACAAACGGGCTGTGTCCCTTTGGTGTGCGGCCAGAGGGGTGGAC 1572  
311 InGlnGlyGlnGlyProTrpGlyThrGlyThrProHisGlyProArg 327  
1573 GG.....AGCCTGGGGCGGTGGACTCCATGGGGGA 1604  
328 LeuGluProAspProGlnHisProGlyAlaTrpLeu.ProLeuLeuSera 344  
1605 CTCGACCGGAC.....CTGTGGCGGCGCGGTGCCICTT 1639  
344 snGlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSerPro 360  
1640 CTAGTCGTCTACTGC.....GACAGCCCC 1662  
361 IleProArg.....CysSerGlyGluSerGluGlnLeuArg 372  
1663 AGGCCAACCATCGGGGCAAGTACTGTCTGGGTGAGAGAAGCGGCACCG 1712  
372 gAlaCysSerGlnAlaProCysProProGluGlnProAspProArgAlaL 389  
1713 CTTCTCAACACAGGATGACTGTCCCTCGCTCCAGGACTTCACAGAG 1762  
389 euGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGlnLeuTrpGln 405  
1763 TGCAGTGTCTGAATTTGACAGCATCCCTTTCCGTGGGAAATTCACAG 1812  
406 TrpGluProPheThrGluValGlnGlySerGlnArgCysGluLeuAsnCy 422  
1813 TGAACAACCTAC...CGGGAGGGGCGGTGAAGGCTGCTCCTCAGAG 1859  
422 sArgProArgGlyPheArgPheTrpValArgHisThrGluLysValGlnA 439  
1860 CCTAGCGGAAGCTTCACTTCTACAGGAGAGGGCGGCGAGCGGTGGTGG 1909  
439 spGlyThrLeuGlyGlnProGlyAlaProAspIleCysValAlaGlyArg 455  
1910 ACGGGACACCTGCCCTCCAGACAGCGGTGGACATTGGCTCAGTGGCGAA 1959  
456 CysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgProAs 472  
1960 TGAACACAGTGGCTGCGACCGAGTCCCTGGGCTCGACCTCGGGGAGGA 2009  
472 pGlyCysGlyValCysGlyCylAspAspSerThrCysArgLeuValSerG 489  
2010 CAAGTCCGAGTGTGGGGGTGACGGCAGTGCCTCGAGACCATCGAGG 2059  
489 lyaAsnLeuThrAspArgGlyGlyProLeuGlyTrpGlnLysIleLeuTrp 505  
2060 GCCTCTTACCCAGCCTCACCTGGGGCGGGGTACAGGATGTCGTCTGG 2109  
506 IleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSe 522  
2110 ATTCCCAAGGCTCCCTCCACATCTTCAATCCAGGATCTGAACCTCTCT 2159  
522 rAsnTrpLeuAlaLeuArgGlyProGlyArgSerIleIleAsnGlyA 539  
2160 CAGTCACTTGGCCCTGAAGGGAGACAGGAGTCCCTGCTGCTCGAGGGC 2209  
539 snTrpAlaValasProProGlySerTrpArgAlaGlyGlyThrValPhe 555  
2210 TGCTGGGACCCCGCCAGCCCGCCAGCTGCTGCTGCTGAGACACCTTT 2259  
556 ArgTrpAsnArgProProArgGluGluGlyLysGlyGlySerLeuSerAl 572  
2260 CAAGTGGAGGGGCGGACAGACAGGTC.....CAGAGCTCGAAGC 2300  
572 aGluGlyProThrThrGlnProValaspValTrpMetIlePheGlnGlu 589  
2301 CCGGGGACCGATTAAATGCAATCTCTCATCGTATGCTGCTGGCCCGGACCG 2350  
589 lyaAsnProGlyValPheTrpGlnTrpValIleSerSer.ProProProIle 605  
2351 AGCTGCTGCTCCCTCGCTACCGCTTC.....AATGCCCATCGCCCGT 2394  
605 eLeuGluAsnProThrProGluPro.....ProValProGlnL 618  
2395 GACTCGCTGCCCC...CCTACTCTGTCATATGCGCCCTGGACCAAGTG 2441  
618 euGlnProGlu..... 621  
2442 CTCGCCCATGTCGACGGCGGTAGCCAGGTGCAGGGGTGGAGTGCAGCA 2491  
622 .....Ile.LeuArgValGluProProLeuAlaProAlaProArgP 635  
2492 ACCAGCTGGACAGCTCGCGGTGCGCCCGCCACTACTGCTGCGCCACAGC 2541  
635 roAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGln.Me 651  
2542 AAGCT.....GCCAAAAGGACGCGCGCTGCACACAGGA 2576  
651 tProAlaProProHisProArgThrProLeuGlySerProAlaAlaTrpT 668  
2577 GCCTTGCCCTCCAGAC.....T 2593  
668 rPlasArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyVal 684  
2594 GGGTGTAGGAAGTGTGCTGCTGAGCCGCGCTGCGATGCGAGGGGTG 2643  
685 TrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGlu..... 699  
2644 CGCAGTACGTGGTGTGTCGACGCGCGGTCTCTGCGCGGAGAGAA 2693  
700 .....LeuAspGluArgSerCysAlaAlaGlyAlaArgProAlaSerP 715  
2694 GCGCTGGACGACAGCGCATGCCGCGAG...CCGCGCCACCTGTACTG. 2739  
715 roGluProCysHisGlyThrProCysProProTrpTrpGluAlaGlyGlu 731  
2740 ..GAGGCTGCGCACGGCCCGCCACTTGCCTCGGAGTGGCAACCCCTGAC 2787  
732 TrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLe 748  
2788 TGGTGTAGTGTACCAAGCTGTGGCGCTGCTCCGCGACCGAGTGGT 2837  
748 uGlnCysArgGlnGluPheGlyGlyGlySerSerValProProGluA 765  
2838 CTTTGTAAAGT.....GCAGATCAACGATCTACTGCTGCCCTCGGC 2881  
765 rGlyGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArg 781  
2882 ACTGCTCTCTCTGACAGCCAGCCATCTACTATGCGATGTAACCTGCGC 2931  
782 LeuCys.....GlyHisTrpGluValGlySerProTrpSerGlnCysSe 796  
2932 CGCTGCCCTCTGCCCGCTGG...GTGACCAAGTGTGGGTGGTGTCTC 2978  
796 rValArgCysGlyArgGlyGlnArgSerArgGlnValArgCysValGlyA 813  
2979 CACACAGTGTGGCTCGGCCAGCAGCAGCAGCAGTGGCTGCACCGCC 3028

the 5' -end sequence/3' -end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2964 BP; 705 A; 862 C; 796 G; 601 T; 0 other;

XX  
XX  
S

```

692 IleSerArgGluSerGlyGluGluLeuAspGluArgSerCysAlaIaGI 708
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      GTGCAGACACATCATGAGAGGCTCCTGAGAGTTACTGTGACTCCAG 637
PT      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
PT      708 yAlaArgProProAlaSerProGluProCysHisGlyThrProCysProP 725
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      638 CATGAAGCGCACCCCGAGGAGGAGCCCTGCAACATCTTCCCTTGCCAG 687
PS      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      725 rOTyTrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro 741
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      688 CTrTCTGGACATCGGGAGTGGTCTGAGTCGACGAAGACCTGTGGCTG 737
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      742 GlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGlyG 758
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      738 GGCATGCAGCACCGCCAGGTCTGTGCGCCAGGTGTACGCCAACCCGAG 787
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      758 ySerSerValProProGluArgCysGlyHisLeuProArgProAsnIleT 775
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      788 CTrGAGGTGACCCCTACCCCTGCCACGACCTGGAGAACTTGAGACCA 837
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      775 hrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluValGlySerPro 791
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      838 CCAGCACCTGCCAACTCAAGATCTGCAGCGAGTGGCAGATCCGACCGAC 887
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      792 TrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerArgGlnVa 808
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      888 TGGACCTCGTCTCGTCCCTCCCTGCGCGCTGGGACAGAGACCGGTGATGT 937
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      808 lArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluCysAlas 825
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      938 GAAGTGTGTGAGCAACATTTGGGGATGTGTTGACGATGAGGAATGC...A 984
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      825 eRgLyProProGlnProProSerArgGluAlaCysAspMetGlyProCys 841
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      985 ACATGAAGCTCGCGCGGAATGACATGAGAACTGGACATGGGACCCCTGT 1034
CC      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC      842 ThrThrAlaTrpPheHisSerAspTrpSerSerLysValSerProGlu 857
XX      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      1035 GCCAAGAGCTGTTCTCACCAGGTGGAGCGAAGAGAGCTCAGCGGAG 1082

```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AAF63448

seq\_documentation\_block:

ID AAF63448 standard; cdna; 3409 BP.

XX AAF63448;

XX 14-MAY-2001 (first entry)

XX Human ADAMTS-10 alternative cdna sequence.

XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;  
 XX tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;  
 KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;  
 KW metastasis; embryogenesis; egg implantation; ADAMTS-10; ss.

XX Homo sapiens.

OS WO200111074-A2.

XX 15-FEB-2001.

XX 03-AUG-2000; 2000WO-US21223.

XX 06-AUG-1999; 99US-0369364.

XX (CLEV-) CLEVELAND CLINIC FOUND.

PA (APTE/) APTE S S.

PA (HURS/) HURSKAINEN T L.

PA (HIRO/) HIROHATA S.

XX Apte SS, Hurskainen TL, Hirohata S;

XX

WPI: 2001-159978/16.  
 P-PSDB; AAB72300.

XX Murine and human 'A Disintegrin-like And Metalloprotease domain with  
 PT thrombospondin type I motifs' proteins and the nucleic acids encoding  
 PT them, useful for treating e.g. tumours, inflammation and arthritis -  
 XX Disclosure; Fig 16; 181pp; English.

XX This invention relates to murine and human ADAMTS-N (A disintegrin-like  
 CC and metalloprotease domain with thrombospondin type I motifs) proteins,  
 CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the  
 CC invention are cDNA sequences encoding the proteins, and antibodies  
 CC specific for the proteins. The nucleic acid sequences and proteins may be  
 CC used in the prevention, diagnosis and treatment of diseases associated  
 CC with inappropriate ADAMTS-N expression. Disorders that may be treated  
 CC using the nucleic acids, proteins and antibodies include, for example,  
 CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos  
 CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage  
 CC in arthritic (both inflammatory and non-inflammatory) disease,  
 CC angiogenesis, tumour growth and metastases, and they may also be used for  
 CC controlling embryogenesis and implantation of fertilised eggs. The  
 CC present sequence represents cDNA encoding human ADAMTS-10.

XX Sequence 3409 BP; 646 A; 1120 C; 1086 G; 557 T; 0 other;

alignment\_scores:  
 Quality: 788.00 Length: 962  
 Ratio: 1.666 Gaps: 46  
 Percent Similarity: 49.168 Percent Identity: 28.067

alignment\_block:  
 US-10-041-770-2 x AAF63448 ..

Align seg 1/1 to: AAF63448 from: 1 to: 3409

4 TrpThrGlyArgProTrpLeuTyrrLeuLeuLeuLeuSerLeuProGI 20  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 538 TGGAAAGGGGGCGCCATGG...TGGCTGGGACCTTGAAGCCACCGCCCTGC 584  
 20 nLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuGluThrProt 37  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 585 CAGACCCCTGGGGAATGAA.....A 604  
 37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 605 CAGAGCGTGGCCAGCCAGGCTGAAGCGATCGTCAAGCGAGCGCTAC 654  
 46 lTrpGlyProTrp..... 50  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 655 GTGAGACCCCTGGTGGCTGACAAGATGATGGCTATCAGCGCG 704  
 51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 705 CCGGATGTGGAGCAGTATGTCTGGCCATCATGAACATTTGTCCAAAC 754  
 67 ArgSerArgThr..... 70  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 755 TTTTCCAGGACTCGAGTCTGGAGACCCGTTACATCTCTGTAACCTGC 804  
 71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 805 CTCATCTCTCTCAGCGAGGACGACGCCACTCTGGAGATCA..... 844  
 85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlyGly 101  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 845 ..CCACCATGCCGGGAAGTCCCTAGACAGCTTCTGTAAGTGGCAGAAAT 892  
 102 ProArgProGlnThrSerProGluThrLeuProLeuTyrrArgThr..... 116  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 893 CCATCTGTGACACGCGCCCATGGCAATGCCATTCATCCAGAGACGGTGTG 942

[illegible]

seq_name:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:AA597171
seq_documentation_block:	
ID	AA597171 standard; cDNA; 2805 BP.
XX	
AC	AA5971716;
XX	
DT	26-FEB-2002 (first entry)
XX	
DE	Human metalloprotease partial DNA sequence #5.
XX	
XX	Human; protease; PCR primer; cytotstatic; immunomodulator; cardiant;
KW	vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquil;
KW	hypertensive; hypoglycemic; neuroleptic; neuroprotective; anabolic;
KW	anorectic; antinflammatory; aspartyl protease; cysteine protease;
KW	metalloprotease; serine protease; cancer; haematopoietic; breast; c
KW	lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;
KW	immune-related disease; cardiovascular disease; neuronal disease;
KW	migraine; sexual dysfunction; mood disorder; attention disorder;
KW	cognition disorder; hypotension; hypertension; psychotic disorder;
KW	dyskinesia; metabolic disorder; inflammatory disorder; ss.
XX	
OS	Homo sapiens.

[illegible]



```

360 ....ProileProArg.....CysSerGlyGluSerGluG 370
    ||| |||:
939 TAACCCCAAGCCTCAGTATGGTGGCTTATCTCTCCAGGTCTAGCGTA 988
    ||| |||:
370 lnLeuArgAlaCysSerGlnAlaProCysProProGluGlnProAspPro 386
    ||| |||:
989 TTTATCAGCTGTGCATATTAACCCCTTGAATGAAATATGCTTGGATT 1038
    ||| |||:
387 ArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGlnLe 403
    ||| |||:
1039 CGGGCTCAACAGTGTGCAGAAATTAACAGCAAAACCTTTCGTGGATGTT 1088
    ||| |||:
403 uTyrGlnTrpGluProPheThrGluValGlnGlnSerGlnArgCysGluL 420
    ||| |||:
1089 CTACCAAGTGAACCCCTATACAAAGTGAAGGAGGAGATCATGATGCAAC 1138
    ||| |||:
420 euAsnCysArgProArgGlyPheArgPheTyrValArgHisThrGluLys 436
    ||| |||:
1139 TGTACTCAAGGCTGAGAACTTGAATTTTGTGCAATGTCCGGCAAA 1188
    ||| |||:
437 ValGlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysValAl 453
    ||| |||:
1189 GTAAAGATGAACTCCTCTCCCAACAAATAATGATTTGTATGA 1238
    ||| |||:
453 aGlyArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgA 470
    ||| |||:
1239 CGGGTTTGTGAACACTAGTGGATGTGATCATGAAGTCTCTAAAGCAG 1288
    ||| |||:
470 rgProAspGlyCysGlyValCysGlyCysAspSerThrCysArgLeu 486
    ||| |||:
1289 TTTTCAGATGCTTGTGGGTTTGAAGGTTGATAATCACTTGCAGATTT 1338
    ||| |||:
487 ValSerGlyAsnLeuThrAspGlyGlyProLeuGlyTyrGlnLysII 503
    ||| |||:
1339 TATAAAGCCTGTACCTCAACAGCATTAAGCAATGAATATTATCCGGT 1388
    ||| |||:
503 eLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgp 520
    ||| |||:
1389 GGTCTCTATTCAGCTGTGGCCGCGCAAGCATCGAAATCCAGGAGCTGCAG 1438
    ||| |||:
520 roSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSerIleIle 536
    ||| |||:
1439 TTTCTCTCAGTTACTCGCAGTTTCAAGCTCAGTCAAAAGTATTACCTC 1488
    ||| |||:
537 AsnGlyAsnTrpAlaValAspProProGlySerTyrArgAlaGlyGlyTh 553
    ||| |||:
1489 ACCGGGCTGTGAGCATCGACTGCTGGGGAGTTCCCTTCGTGGGAC 1538
    ||| |||:
553 rValPheArgTyr.....AsnArgProProArgGluGluGlyL 566
    ||| |||:
1539 CACGTTTGAATACAGCGCTCTTCAACCGCCG..... 1572
    ||| |||:
566 ysGlyGluSerLeuSerAlaGluGlyProThrThrGlnProValAspVal 582
    ||| |||:
1573 ....GAACGCTGTACGCGCAGGCGCCCAAAATGAGACGCTGCTTT 1617
    ||| |||:
583 TyrMetIlePheGlnGluAsnProGlyValPheTyrGlnTyrValII 599
    ||| |||:
1618 GAAATTCGTATGCAAGCAAAATCCAGGATAGCTTGGAGATGATGCACT 1657
    ||| |||:
599 eSerSerProProIleLeuGluAsnProThrProGluProProValp 616
    ||| |||:
1668 T..... 1668
    ||| |||:
616 roGlnLeuGlnProGluIleLeuArgValGluProProLeuAlaProAla 632
    ||| |||:
1669 .....CCCAAGGTCATGATGAACTCCACGACCCACAAAGA 1707
    ||| |||:
633 ProArgProAlaArgThrProGlyThrLeuGlnArgGlnValArgIlePr 649
    ||| |||:
1708 CCT..... 1710
    ||| |||:
649 oGlnMetProAlaProProHisProArgThrProLeuGlySerProAlaA 666

```

```

1711 .....G 1711
    ||| |||:
666 laTyrTrpIlyArgValGlyHisSerAlaCysSerAlaSerCysGlyLys 682
    ||| |||:
1712 CCTATACCTGGAGTATCGTCAGTCAGAGTGTCTCGTCTCTGTGTGGA 1761
    ||| |||:
683 GlyValTrpArgProIlePheLeuCysIleSerArgGluSerGlyGluL 699
    ||| |||:
1762 GGTTCATATAATGTAAAGGCCATTGCTTG...CGAGATCAAAATACTCA 1808
    ||| |||:
699 uLeuAspGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProG 716
    ||| |||:
1809 AGTCATATCTCTCATTCGTCAGTGCAAAACCAAGCAGTAACTAGAGCCA 1858
    ||| |||:
716 luProCysHisGlyThrProCysProProTyrTrpGluAlaGlyGluTrp 732
    ||| |||:
1859 AAATCTGCAACGCTTCTCTCCCGGCTTACTGATGCCAGGTGAATGG 1908
    ||| |||:
733 ThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuG 749
    ||| |||:
1909 AGTACATGACGAAAGCCTGCTGTGAGGCCAGCAGAGCCGAAAGATCCA 1958
    ||| |||:
749 nCysArgGlnGluPheGlyGlyGlySerSerValProProGluArgC 766
    ||| |||:
1959 GTGTGTGCAAAAGAACCCCTTCCAAAGGAGGAGCAGTGTTCATTCCTC 2008
    ||| |||:
766 ysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeu 782
    ||| |||:
2009 TCTGTCCAGTGAACACCCATCAGTCCAGCCTGCAACAGCATGCC 2058
    ||| |||:
783 Cys...GlyHisTrpGluValGlySerProTrpSerGlnCysSerValar 798
    ||| |||:
2059 TGCCCTCCCAATGGAGCTTGA...CCTGTCTCAGTGTTCCAAGAC 2105
    ||| |||:
798 gCysGlyArgGly..... 802
    ||| |||:
2106 CTGTGACGAGGGGTGAGGAGCGTGAACCTCTCTGCAAGGGCTCGCG 2155
    ||| |||:
803 .....GlnArgSerArgGlnValArgCysValGlyAsnAsn 814
    ||| |||:
2156 CAGAAACCCCTCCCGAGAGGAGCGTGAACCTCTCTGCAAGGGCTCTGCC 2205
    ||| |||:
815 GlyAspGluValSerGlnGluCysAlaSerGlyProProGlnProPr 831
    ||| |||:
2206 GCAGAAACCCCTCCCGAGAGCGAGTACCAGT...CTCCCGACGACCTGA 2252
    ||| |||:
831 oSerArgGluAlaCysAspMetGlyProCys.....ThrThra 844
    ||| |||:
2253 GCTGACGAGGGGTGTGCTTGGACGATGCCCAAGAACAGCGGCTAC 2302
    ||| |||:
844 laTrpPheHisSerAspTrpSer 851
    ||| |||:
2303 AGTGGTCTGCTTCTCTGTCGAGC 2325
    ||| |||:

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC75588

seq\_documentation\_block:

ID AAC75588 standard; CDNA; 6505 BP.

XX

AC AAC75588;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF1143 polynucleotide sequence SEQ ID NO:2285.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antithyroid; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 cholesterol ester storage; systemic lupus erythematosus; infection;  
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WO2000058473-A2.

XX PD 05-OCT-2000.

XX PF 31-MAR-2000; 2000WO-US08621.

XX PR 31-MAR-1999; 99US-0127607.

XX PR 02-APR-1999; 99US-0127636.

XX PR 05-APR-1999; 99US-0127728.

XX PR 30-MAR-2000; 2000US-0540763.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX DR WPI; 2000-602362/57.

XX DR P-PSDB; AAB41379.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 5; Page 1674-1678; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human OREF open reading frames 1 to 3161. The OREF  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipapillary; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antihypertensive; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an OREF-associated disorder. The  
 CC nucleic acids can be used to express OREF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 6505 BP; 1154 A; 2202 C; 2005 G; 1141 T; 3 other;

alignment\_scores:

Quality:	757.50	Length:	963
Ratio:	1.698	Gaps:	42
Percent Similarity:	46.314	Percent Identity:	27.207

alignment\_block:

US-10-041-770-2 x AAC75588 ..

Align seg 1/1 to: AAC75588 from: 1 to: 6505

47 TrpGlyProTrp.....

||| |||||

740 TGGAGACCCCTGGTAGCTGATGCCAAATGGTGGAGTACCAGGACAG 789

51 .....ValGlnTrpAlaSerC 56  
 790 CCGCAGGTGAGAGCTATGTGCTGACCATGATGAACATGTGGTGGCT 839  
 56 ySerGlnProCysGlyValGlyVal..... 64  
 840 GTTCATGACCCAGCATGGGAACCCCATCCACATCACCATTGTGGCC 889  
 65 .....GlnArgArgSerArgThrCysGlnLeuProThrValG1 77  
 890 TGGTCTGTGGAAGATGAGGAGGAGGACCTAAAGATCAGCAGCATGCA 939  
 77 nLeuHisPro..... 80  
 940 GACACACCCCTGAAGAGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAA 989  
 81 ..SerLeuProLeuProProArgProProArgHisProGluAlaLeuLeu 96  
 990 GGGGATGCCATCCCTGCCACCATG..... 1015  
 97 ProArgGlyGlnGlyProArgProGlnThrSerProGlnThrLeuProLe 113  
 1016 .....ACACTGCCATCCTCTCACCAGAA..... 1039  
 113 UTyArgThrGlnSerArgGlyArgGlyProLeuArgGlyProAlas 130  
 1040 ....AGGACCTGTGTGCAGCATGACCGGCCCTGTGAGACCCCTGGGACT 1085  
 130 erHisLeuGlyArgGluGluThrGlnGluLeuArgAlaAlaArgSer 146  
 1086 GTCCCATGTGGGGCATGTGCCAGCCGCCAGCTGCAGCATCAACG 1135  
 147 ArgLeuArgAspProIleLysProGlyMetPheGlyTyArgValPr 163  
 1136 AGGACA..... 1141  
 163 oPheAlaLeuProLeuHisArgAsnArgArgHisProArgSerProProA 180  
 1142 .....CGGGCCTGCCCTTTCATCATGTCTCCACAGCTCTGTATC 1172  
 180 rGSerGluLeuSerLeuLeuSerArgGlyGluGluAlaIleProSer 196  
 1173 CGAGCTCGGCACAGTTTGGCATTCACATGACGGAAGCGGCAATGACT 1222  
 197 ProThrProArgAlaGluProPheSerAla.....AsnGlySerProG1 211  
 1223 GTGAGCCGTTGGGAACGACCTTTCATCATGTCTCCACAGCTCTGTATC 1272  
 211 nThrGluLeuPro....ProThrGluLeuSerValHisThrProSerProG 227  
 1273 GACGCGCTCCCTCCCTCACCTGGTCCGCTGCGCCGCCAGTATATCACCAG 1322  
 227 InAlaGluProLeuSerProGluThrAlaGlnThrGluValAlaProArg 243  
 1323 GTTCTTGACCGTGGGTGGGGCTGTGCTGGACGACCTCTCTGCCAAGG 1372  
 244 ThrArgProAlaProLeuArgHis.....Pr 253  
 1373 ACATTATCGACTTCCCTCGGTGCCACCTGGCGTCTCTATGATGTAAGC 1422  
 253 oArgAlaGlnAlaSerGlyThrGluProProSerPro..... 265  
 1423 CACAGTGGCGCTCCAGTACGGGGCTTACTTGTCTTGTGGAGGACAT 1472  
 266 .....ThrHisSerLeuGlyGlyGlyPhePheArgAla 277  
 1473 GGATAATGTCTGCCACACACTCTGTGTCTGTGGGA..... 1510  
 278 SerProGlnProArgProSer....SerGlnGlyTrpAlaSerProG1 293  
 1511 .....CACCTGTCTCAAGCTGGATGAGCTGTGGAGCGGACCCGG 1554  
 293 nValAlaGlyArg.....ArgProAspProPheP 303

```

1555 TGTG...GGGAGATAAGTGTGTCTCACTGGGGAGTCCGTACCCGTGGG 1601
|||||
303 roSerValProArgGlyArgGlyGlnGlnGlyProTyrP... 317
|||||
1602 CTTCCGGCCGAGG... 1627
|||||
318 ...GlyThrGlyGlyThrProHisGlyProArgLeuGluProAspProG1 333
|||||
1628 GCTGTGCTGGCTGGAGCCCT...GGTCCA... 1654
|||||
333 nHisProGlyAlaTyrProLeuProLeuLeuSerAsnGlyProHisAlaSerS 350
|||||
1655 ... 1670
|||||
350 exLeuTyrSerLeuPheAlaProSer...Ser...ProIleProA 363
|||||
1671 TGGCATGGGCTACAGAGCGCGGAGCGAGTGCAGCGAGCTACGCCCA 1720
|||||
363 rg... 374
|||||
1721 ATACAAAGGCAGATACTGTGTGGGTGAGCGCAAGCGCTTCGGCTCTGC 1770
|||||
375 SerGlnAlaProCysProProGluGlnProAspProArgAlaLeuGlnCy 391
|||||
1771 AACCTGCAGGCGCTGCCCTGCTGGCGGCCCTCTCCCGCAGTCCAGTG 1820
|||||
391 sAlaAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrGlnTyrGluP 408
|||||
1821 CAGCCACTTTCAGCTATGCTCTACAAGGGCAGCTGCACACATGGGTGC 1870
|||||
408 roPheThrGluValGlnGlnSerGlnArgCysGluLeuAsnCysArgPro 424
|||||
1871 CGGTG...GTCATGACGTGAACCCCTGGAGCTGCACCTGCCGCGCC 1914
|||||
425 ArgGlyPheArgPheTyrValArgHisThrGluLysValGlnAspGlyTh 441
|||||
1915 GCGATGACTACTTTCGCAAGAGCTGCGGACGCGGTGTCATGGGCAC 1964
|||||
441 rLeuCysGlnPro...GlyAlaProAspIleCysValAlaGlyArgC 456
|||||
1965 CCCTGCTACCAAGTCCGAGCCGAGCGGACCTGCATCAACGGCATCT 2014
|||||
456 ysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgProAsp 472
|||||
2015 GTAAAGACGTGGCTGTGACTTCGAGATGACTCCGGTCTATGAGGAC 2064
|||||
473 GlyCysGlyValCysGlyGlyAspSerThrCysArgLeuValSerGl 489
|||||
2065 CGCTGTGTGTGTGCCACGCAACGCTCCACTGCCACACCGTGGCGG 2114
|||||
489 yAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLysIleLeuTrpI 506
|||||
2115 GACCTTCNAGAGGAGCCGAGGCTCWTGGGTATGGATGTGGGCTGA 2164
|||||
506 leProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSer 522
|||||
2165 TCCACGCGGCGCAGCGAGATCCGATCCAGAGGTTCGCGAGGCTGCC 2214
|||||
523 AsnTyrLeuAlaLeuArgGlyProGly...GlyArgSerIleIleAsnGl 538
|||||
2215 AACTCTCTGGCACTGGCGGAGGAGGACCCGAGAGACTTCTCTCAATGG 2264
|||||
538 yAsnTyrAlaValAspProProGlySerTyrArgAlaGlyGlyThrValP 555
|||||
2265 TGGCTGACCATCCAGTGAACGGGACTTACCAGGTGGCAGGACCACT 2314
|||||
555 heArgTyrAsnArgProProArgGluGluGlyLysGlyGluSerLeuSer 571
|||||
2315 TCATATACGACGC...AGGGCAACTGGGAGAACCTCAGC 2352
|||||
572 AlaGluGlyProThrThrGlnProValAspValTyrMetIlePheGlnGl 588
|||||

```

```

2353 TCCCGGGTCCCAACCAAGGAGCGCTGTCTGGATC...CAGGTGCTGCCTC 2399
|||||
588 uGluAsnProGlyValPheTyrGlnTyrValIleSerSerProProI 605
|||||
2400 CCGTGGCCCAAGCGGGGGAGAGCGGAGTCCCGAGCCCGAGCACCC 2449
|||||
605 leLeuGluAsnProThrProGluProValProGlnLeuGlnProGlu 621
|||||
2450 TCCATGSCAGGTCTCGTCTGGAGGAGTGGCTTGGTTCAGTC... 2493
|||||
622 IleLeuArgValGluProProLeuAlaProAlaProArgProAlaArgTh 638
|||||
2494 ...ACAGAGCTGGCTGTGAGCCAGGCCCT...CCTGCTCGGC 2531
|||||
638 rProGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProP 655
|||||
2532 CTCTACCTCAGTTCCCATCTTTAAATGCCCCAATCTTGTAGCTGCAG 2581
|||||
655 roHis... 656
|||||
2582 TTCACAGAGGTGGTGGGGTCAAGCTCCTTTAGGACTGGGTGGATGGAGA 2631
|||||
657 ...ProArgThrProLeu... 661
|||||
2632 AGACACCTTGTGCTCATGGGCCCGCCCTGCCACCCAGCTGTGTTC 2681
|||||
662 ...GlySerProAlaAlaTyrTrpLysArgVal... 671
|||||
2682 GGAGAGAACCTTGGGTGCTACAGTACAGCATCCACAGGAGGAGCAG 2731
|||||
672 ...GlyHis... 673
|||||
2732 GTGGCCAGGAGAGGTCCGCGGCCGCTGTCTCTGGCATTATGGGCC 2781
|||||
674 ...SerAlaCysSerAlaSerCysGlyValGlyValTyrArgProIlePh 689
|||||
2782 TGGACCAAGTGCACAGTCCACCTGGCGGAGAGTGTGCAGAGCGAGAATGT 2831
|||||
689 eLeuCysIleSerArgGluSerGlyGluLeuAspGluArgSerCysA 706
|||||
2832 GTACTGTGGAGCGGAGGAGG...CCGTGGACGAGGAGCAGCTGTG 2878
|||||
706 laAlaGlyAlaArgProProAlaSerProGluProCysHisGlyThrPro 722
|||||
2879 ACCCCTGGCGCGCTGTATGACCAACAGAGGAAGTGCAGCGAGCAGCC 2928
|||||
723 CysProProTyrTrpGluAlaGlyGluTyrThrSerCysSerArgSerCy 739
|||||
2929 TGCCTGCGCAGGTGGTGGCAGGTGAGTGGCAGCTGTCTCCAGCTCTG 2978
|||||
739 sGlyPro...GlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPhe 755
|||||
2979 CGGCGCTGGGCGCTCTCCCGCGGCGCTGTCTGTGATCCGAGCGTGG 3028
|||||
755 ly...GlyGlyGlySerSerValProProGluArgCysGlyHisLeuPro 770
|||||
3029 GGCTGGATGACGAGAGCGCTGAGCCCGCTGTGACACCTTCCCTCCC 3078
|||||
771 ArgProAsnIleThrGlnSerCysGlnLeuArgLeu...Cys...GlyH1 785
|||||
3079 CGGCCCCCTACTGAACCCCTTGCAACCGCCATGTACCTGTCCCGGCCAC 3128
|||||
785 sTrpGluValGlySerProTyrSerGlnCysSerValArgCysGlyArg 802
|||||
3129 CTGGCTGTGGGGAAC...TGGTCTCAGTGTCTGACATGTGGGAGG 3175
|||||
802 lyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluVal 818
|||||
3176 GCACACGCGCGGAATGTCTCTGCACCAATGACCGGCTGTCCCTGT 3225
|||||
819 serGluGlnGluCysAlaSerGlyProProGlnProProSerArgGluAl 835
|||||
3226 GACGAGGCGCCAG...CAGCCAGCGCAGCAAGTCAC 3257
|||||

```

```
835 acysaspMetGlyProCysThr.ThrAlaTrpPheHisSerAspTrpSer 851
:||||:||||:||||:||||:||||:||||:||||:||||:
3258 CTGCTCTGCGCACTCTGTGGTGGCCCTGGGCACAC.....TGGGCC 3301

852 SerLysVal...SerProGluProProAlaIleSer 862
||||:||||:||||:||||:||||:||||:||||:||||:
3302 CTGAAGGCTCAGGAGGCGGCTCTCCAGCCACGAGC 3337
```

OM of: US-10-041-770-2 to: GenEmbl.\* out\_format : pfs  
Date: Jul 23, 2002 10:28 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=framet\_p2h.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US10041770/runat\_22072002\_153135\_2679/app\_query.fasta\_1.947  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US10041770 -CGN1\_1\_5038  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPY -WAIT -THREADS=1

## Search information block:

Query: US-10-041-770-2  
Query length: 877  
Database: GenEmbl.\*  
Database sequences: 187333701  
Database length: 187333701  
Search time (sec): 2782.860000

## score\_list:

Std	Orig	ZScore	Escore	Len	Documentation
gb_pat:AX342636	+	4769.00	1381.44	4230	AX342636 Sequence 33 from Pat
gb_pr:AL356356	+	3976.50	1139.91	3.9e-55	AL356356 Human DNA sequence
gb_pat:AC092479	+	2670.50	769.90	1.1e-34	AC092479 Mus musculus clone
gb_pat:AC093317	+	2651.50	762.67	4.1e-34	AC093317 Mus musculus clone
gb_pat:AC053497	+	2070.50	597.80	6.2e-25	AC053497 Homo sapiens chrom
gb_pat:AC053497	+	2014.50	581.90	4.8e-24	AC053497 Homo sapiens chrom
gb_pr:AF211974	+	1231.00	379.63	8.8e-13	AF211974 Homo sapiens clone PH
gb_pr:BC008840	+	935.50	294.52	4.8e-08	BC008840 Homo sapiens, hypoth
gb_pat:ES5265	+	807.50	255.40	7.3e-06	ES5265 Novel metalloprotease a
gb_pat:AX128433	+	806.50	254.87	7.8e-06	AX128433 Sequence 15 from Pat
gb_pat:AX1319851	+	797.50	254.43	3.3e-06	AX1319851 Sequence 15 from Pat
gb_pat:ES5273	+	796.50	254.14	8.6e-06	ES5273 Novel metalloprotease a
gb_pr:AF163762	+	791.50	252.61	1.0e-05	AF163762 Homo sapiens zinc met
gb_pr:AK023772	+	789.50	252.64	1.0e-05	AK023772 Homo sapiens cDNA FLJ
gb_pat:AX327754	+	784.00	250.84	1.3e-05	AX327754 Sequence 10 from Pat
gb_pat:AX327752	+	784.00	250.57	1.4e-05	AX327752 Sequence 18 from Pat
gb_pat:AX319854	+	780.00	250.19	1.4e-05	AX319854 Sequence 11 from Pat
gb_pat:AX327755	+	759.50	243.43	3.4e-05	AX327755 Sequence 1 from Pat
gb_pat:AX111571	+	757.00	243.07	3.6e-05	AX111571 Sequence 1 from Pat
gb_pat:AX327747	+	757.00	243.06	3.6e-05	AX327747 Sequence 3 from Pat
gb_pat:AX338539	+	757.00	242.80	3.7e-05	AX338539 Sequence 3 from Pat
gb_pat:AX327745	+	757.00	242.80	3.7e-05	AX327745 Sequence 1 from Pat
gb_pat:AX338537	+	757.00	242.80	3.7e-05	AX338537 Sequence 1 from Pat
gb_pat:AX338537	+	757.00	242.80	3.7e-05	AX338537 Sequence 16 from Pat
gb_pat:AX319852	+	738.50	241.47	4.4e-05	AX319852 Sequence 1 from Pat
gb_ro:AF314171	+	738.50	236.55	4.2e-05	AF314171 Mus musculus papilin
gb_in:AF078161	+	731.00	230.25	0.0002	AF078161 Manduca sexta lacunin
gb_in:AF205357	+	722.50	229.77	0.0002	AF205357 Drosophila melanogast
gb_pat:AX111573	+	714.50	232.52	0.0001	AX111573 Sequence 3 from Pat
gb_pat:AX111573	+	714.50	232.52	0.0001	AX111573 Sequence 3 from Pat
gb_pat:AX111573	+	699.50	212.24	0.0019	AX111573 Drosophila melanog
gb_in:AC014412	+	699.50	208.95	0.0028	AC014412 Drosophila melanog
gb_in:AE009213	+	699.50	207.76	0.0033	AE009213 Drosophila melanog
gb_in:AE003765	+	699.50	226.34	0.0003	AE003765 Drosophila melanog
gb_pat:AX319861	+	695.50	223.47	0.0004	AX319861 Sequence 25 from Pat
gb_in:AF043121	+	695.50	223.47	0.0004	AF043121 Haemochromis concolor
gb_pr:AB011177	+	676.00	219.01	0.0008	AB011177 Homo sapiens mRNA for
gb_pr:HS250725	+	666.00	215.16	0.0013	HS250725 Homo sapiens mRNA for
gb_pat:AX319860	+	636.00	209.08	0.0028	AX319860 Sequence 24 from Pat
gb_pat:AX342635	+	626.50	206.39	0.0039	AX342635 Sequence 32 from Pat
gb_pat:AX342638	+	601.00	196.87	0.0089	AX342638 Sequence 35 from Pat
gb_pr:HS305314	+	598.50	200.61	0.0133	HS305314 Homo sapiens mRNA for
gb_pr:HS305314	+	598.50	200.61	0.0133	HS305314 Homo sapiens mRNA for
gb_pr:HS305314	+	596.00	196.78	0.0135	HS305314 Homo sapiens mRNA for

gb\_pr:AY055376 + 593.50 195.04 0.0168 4597 i AY055376 Homo sapiens von W  
gb\_pr:HS420810 + 590.50 194.50 0.0180 4278 i A1420810 Homo sapiens mRNA  
gb\_pr:AF414401 + 590.50 194.23 0.0187 4550 i AF414401 Homo sapiens ADAMT  
gb\_pr:AB056988 + 590.50 193.86 0.0196 4950 i AB056988 Homo sapiens vWF-C  
gb\_pr:AB037733 + 589.00 193.27 0.0211 5139 i AB037733 Homo sapiens mRNA

seq\_name: gb\_pat:AX342636

## seq\_documentation\_block:

LOCUS AX342636 4230 bp DNA linear PAT 12-JAN-2002  
DEFINITION Sequence 33 from Patent WO0198468.  
ACCESSION AX342636  
VERSION AX342636.1 GI:18152033

## KEYWORDS

human.

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (sites)

Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,

Tribouley, C.M., Deleage, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,

Hafalia, A., Khan, F.A., Walla, N.K., Yao, M.G., Lu, D.A., Patterson, C.,

Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R.

Patent: WO 0198468-A 33 27-DEC-2001;

Incyte Genomics, Inc. (US)

## FEATURES

Location/Qualifiers

1..4230

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/note="Incyte ID No: 7604035CB1"

BASE COUNT 810 a 1411 c 1210 g 799 t

## ORIGIN

## alignment\_scores:

Quality: 4769.00 Length: 857

Ratio: 5.584 Gaps: 0

Percent Similarity: 99.650 Percent Identity: 99.650

## alignment\_block:

US-10-041-770-2 x AX342636

Align seg 1/1 to: AX342636 from: 1 to: 4230

1 MetGluAsnTrpThrGlyArgProTrpLeuTyLeuLeuLeuLeuLeu 17

|||||

228 ATGGAGAACTGGACGGGAGGCGCCGCGTGTATCTGCTGCTCTGTC 277

|||||

17 rLeuProGlnLeuCysLeuAspGlnValLeuSerGlyHisSerLeuG 34

|||||

278 CCTCCCTCAGCTCTGCTTGGATCAGGAGGTGTGTCCGACACTCTCTC 327

|||||

34 lnThrProThrGluGluGlnGlyProGluGlyValTrpGlyProTrp 50

|||||

328 AGACACCTACAGAGGAGGCGCCGAGGCGCCGAGGTCGCGGACCTTGG 377

|||||

51 ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 67

|||||

378 GTCCAGTGGCGCTCTGCTCCAGCCCTCCGGGTGGGTGCAGCGCAG 427

|||||

67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuPro 84

|||||

428 GAGCCGAGATGTGACCTCCCTACAGTGCAGCTCCACCGAGTCTGCC 477

|||||

84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100

|||||

478 TCCTCCCGCGCGCCCAACATCCAGAGCCCTCTCCCGCGGCGCAG 527

|||||

101 GlyProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGln 117

|||||

528 GGTCCCGAGACCCAGACTTCTCCAGAAACCTCCCTTGTACAGGACACA 577

|||||

117 nSerArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyA 134

578 GTCTGGGGAAGGGGTGGCCCACTTCGAGTCCCGCTTCCACCTAGGGA 627  
 134 rGluGluThrGlnGluLeuAlaArgArgSerArgLeuArgasp 150  
 628 GAGAGGAGCCAGGAGATTTCAGAGGCGCCAGGAGGTCCCGCTTCGAGAC 677  
 151 ProIleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuPr 167  
 678 CCATCAGCCAGGATGTTCGGTTATGGGAGAGTGCCCTTTGCATTGCC 727  
 167 OLeuHisArgAsnArgHisProArgHisProArgSerProProArgSerGluLeu 184  
 728 ACTGCACCGCAACCCAGCAGCCCTCGAGGCGCCAGCCAGATCTGAGCTGT 777  
 184 erLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProArg 200  
 778 CCTGATCTCTCTAGAGGGGAAGAGCTATTCCGTCCCTTACTCCCAAGA 827  
 201 AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProTh 217  
 828 GCAGAGCCATTCTCGCAACAGCAGCCGCCCAACTGAGCTCCCTCCAC 877  
 217 rGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG 234  
 878 AGAAGCTGTCTCCACACCCCTATCCCTCCCAAGCAGAACCTCTAAGCCCTG 927  
 234 LuThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArg 250  
 928 AAATCTCTCAGACAGAGGTGGCCCCAGAACCCAGAGCCCTGCCCTACGG 977  
 251 HisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrH 267  
 978 CATCACCCAGAGCCAGGCCCTCTGGCACAGAGCCGCCCTCACCCACGCA 1027  
 267 sSerLeuGlyGluGlyPhePheArgAlaSerProGlnProArgArgP 284  
 1028 CTCCTTAGGAGAAGGTGGCTTCTTCGGTGCATCCCTCAGCCACAGAGC 1077  
 284 roSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArgArgProasp 300  
 1078 CAAGTTCACAGGTTTGGGCCAGTCCCCAGGTAGCAGGGAGAGCGCCTGAT 1127  
 301 ProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProTr 317  
 1128 CTTTTCCTTCGGTCCCTCGGGCGGAGGCCAGCAGGCGCAAGSGCCTTG 1177  
 317 pGlyThrGlyThrProHisGlyProArgLeuGluProAspProGlnH 334  
 1178 GGAACCGGGGGGACCTCTCAGGGCGCCCGCTGGAGCGCTGACCCCTCAGC 1227  
 334 isProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSerSer 350  
 1228 ACCGGGCGCTGGCTGCCCCCTGCTGAGCAAGCGGCCCATGCCAGTCC 1277  
 351 LeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyGl 367  
 1278 CTCTGAGCGCTCTTCTGCCAGTAGCCCTATTCCAAAGATGTTCTGGGGA 1327  
 367 uSerGluGlnLeuArgAlaCysSerGlnAlaProCysProProGluGlnP 384  
 1328 GAGTGAACAGCTAAGAGCTGAGCAAGCGCCCTGCCCTCCCTGAGCAGC 1377  
 384 roAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMet 400  
 1378 CAGACCCCGGGCCCTGCGAGTGGCGAGCCTTTAACTCCAGCAATTCATG 1427  
 401 GlyGlnLeuTrpGlnTrpGluProPheThrGluValGlnGlnGlySerGlnAr 417  
 1428 GCCAGCTGTATCAGTGGAGCGCCTTCACTGAAGTCCAGGGCTCCAGCG 1477  
 417 gCysGluLeuAsnCysArgProArgGlyPheArgPheTyrValArgHisT 434

1478 CTGTGAACCTGAACCTGCCGGCCCGCTTCCGCTTCTATGTCCGTCA 1527  
 434 hrGluLysValGlnAsnGlyThrLeuCysGlnProGlyAlaProAspIle 450  
 1528 CTGAAAAGGTCCAGATGGAGCCCTGTGTGAGCTGGAGCCCTTGACATC 1577  
 451 CysValAlaGlyArgCysLeuSerProGlyCysAspGlyIleLeuGlySe 467  
 1578 TGTGTGCTGGACGCTGTCTGAGCCCGCGCTGTATGGGATCCTTGCTC 1627  
 467 rGlyArgArgProAspGlyCysGlyValCysGlyGlyAspAspSerThrC 484  
 1628 TGGCAGCGCTCTGTATGGCTGTGGAGTCTGTGGGGTGTATCTACCT 1677  
 484 ysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyr 500  
 1678 GTCCGCTTGTTCGGGGAACCTCACTGACCGAGGGGCCCTTGGGCTAT 1727  
 501 GlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGl 517  
 1728 CAGAAGATCTTGTGATTCAGCGGGAGCCTTGGGCTCCAGATTCGCCA 1777  
 517 nLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgS 534  
 1778 GCTCCGGCTAGCTCCAACTACCTGGCACCTTCGTGGGCCCTGGGGCGCGT 1827  
 534 erIleIleAsnGlyAsnTrpAlaValAspProProGlySerTyrArgAla 550  
 1828 CCATCATCAATGGGAACCTGGGCTGTGGATCCCCCTGGGTCTCTACAGGCC 1877  
 551 GlyGlyThrValPheArgTyrAsnArgProProArgGluGluGlyLysGl 567  
 1878 GGCGGGACCGTCTTTCGATATACCGTCTCCAGGAGGAGGAGGCAAGG 1927  
 567 yGluSerLeuSerAlaGluGlyProThrThrGlnProValaspValTyrM 584  
 1928 GGAGAGTGTGCGCTGAAAGGCCCCACACCCAGCCCTGTGATGTCTATA 1977  
 584 erIlePheGlnGluAsnProGlyValPheTyrGlnTyrValIleSer 600  
 1978 TGATCTTTCAGAGGAAACCCAGCGCTTTTATCAGTATGTCATCTCT 2027  
 601 SerProProIleLeuGluAsnProThrProGluProProValProGl 617  
 2028 TCACCTCTCCAATCTTGAGAACCCACCCAGCCCGCTGTCCCCCA 2077  
 617 nLeuGlnProGluIleLeuArgValGluProProLeuAlaProAlaProA 634  
 2078 GCTTCAGCCGGAGATTCTGAGGGTGGAGCCCGCCACTTGCTCCCGCACCCC 2127  
 634 rgProAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGln 650  
 2128 GCCCAGCCCGGACCCAGGCCACCTCCAGCGTCAGGTCCGGATCCCCCAG 2177  
 651 MetProAlaProProHisProArgThrProLeuGlySerProAlaAlaTy 667  
 2178 ATGCCCGCCCGCCCATCCAGGAGACCCCTGGGGTCTCCAGCTGGTA 2227  
 667 rTriPlysArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyV 684  
 2228 CTGGAACAGAGTGGGACACTCTGCATCTCAGCGTCTCGGGGAAAGGTG 2277  
 684 alTrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeu 700  
 2278 TCTGGCGCCCATTTCTCTGTCATCTCCCGTGTGAGTCGGGAGAGGAATG 2327  
 701 AspGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluPr 717  
 2328 GATGACCGAGCTGTCCGCGGGTGCAGGCCCCAGCCCTCCCTGAAAC 2377  
 717 oCysHisGlyThrProCysProProTyrTrpGluAlaGlyGluTrpThrS 734  
 2378 CTGCCACGCGCACCCCATGCCCCCATACTGGGAGGCTGGCGAGTGGACAT 2427

```

734 erCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCys 750
|||||
2428 CCTGCACCCGCTCTGTGGCCGCGCACCCAGCAGCCAGCTGCAGTGC 2477
751 ArgGlnGluPheGlyGlyGlySerValProGlnArgCysG1 767
|||||
2478 CGGAGAAATTTGGGGGGGTGGCTCTCGGTGCCCGGAGCGCTGG 2527
767 yHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysG 784
|||||
2528 ACATCTCCCGCGCCCAACATCACCCAGTCTGCCAGCTCGCGCTCTGTG 2577
784 lyHisTrpGluValGlySerProTrpSerGlnCysSerValArgCysGly 800
|||||
2578 GCCATTGGGAAGTTGGCTCTCTCTTGGAGCCAGTCTCCGTGGGTGGCGC 2627
801 ArgGlyGlnArgSerArgGlnValArgCysValGlyAsnGlyAspG1 817
|||||
2628 CGGGCCAGAGAACCGCAGGTTCGCTGTGGGAACAACAGTGATGA 2677
817 uValSerGluGlnCysAlaSerGlyProGlnProProSerArgG 834
|||||
2678 AGTGAGCGAGCAGGAGTGTGGTCAGGCCCCACACGCCCCAGCAGAG 2727
834 luAlaCysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTrp 850
|||||
2728 AGGCCTGTGACATGGGCCCTGTACTACTGCTGCTGTTCCACAGCGACTGG 2777
851 SerSerIysValSerProGlu 857
|||||
2778 AGCTCCAAAGTCTCAGCGGAG 2798

seq_name: gb_pr:AL356356

seq documentation block:
LOCUS AL356356 176550 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone Rp11-54A4 on chromosome 1, complete
sequence.
ACCESSION AL356356
VERSION AL356356.17 GI:18476570
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Heath, P.
Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
Cloned from request@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:18250748.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; SW;
SWISSPROT; Tr; TREMBL; Wp; WormPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

```

```

Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
Rp11-54A4 is from the library RPI-11.1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-54A4. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone Rp11-54A4 is at 1 in this sequence. The
true left end of clone Rp11-363122 is at 174551 in this sequence.
FEATURES
    Location/Qualifiers
        1..176550
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="Rp11-54A4"
            /clone_lib="RPI-11.1"
        misc_feature
            50307..50323
                /note="Single clone region. Reads generated from a
                transposon library derived from a single pUC clone.
                Restriction digest data confirm the assembly."
        misc_feature
            123022..123038
                /note="Sequence from uni-directional primer reads and dGTP
                big dye terminator reads only."
BASE COUNT 45999 a 42218 c 42262 g 46071 t
ORIGIN
alignment_scores:
    Quality: 3976.50      Length: 2064
    Ratio: 4.587          Gaps: 19
    Percent Similarity: 42.006      Percent Identity: 41.860
alignment_block:
US-10-041-770-2 x AL356356 ..
Align seg 1/1 to: AL356356 from: 1 to: 176550
7 ArgProTrpLeuTyLeuLeuLeuLeuLeuSerLeuProGlnLeuCysLe 23
|||||
79803 AGGCCCTGGCTGTATCTGCTGCTTCTCTCTCCCTCCCTCAGCTGCT 79852
23 uAspGlnGlu..... 26
|||||
79853 GGATCAGGAGGTGAGTCTGGACAAGTGAGCAGCTGCAGCCTGCCTGCCA 79902
26 ..... 26
79903 CCCTTTCATCTCCCTAGGCTCCACCCCACTCTCCAGGCCACGCCTC 79952
26 ..... 26
79953 CATTCCCCACAGCCACCCACCTCCCTGCCAATCCCGAGTTCCCTGCC 80002
26 ..... 26
80003 TCATACACCAAGAGCGGAGGTGTAGTTCTCCCTCTGCTGTAATGTG 80052
26 ..... 26
80053 ACCTTGAGCTGTAGCAGACCTCCGTGAGCCCTCAGTTGGTCTACATGGAC 80102
26 ..... 26
80103 ACTCTGACAGTTCCCTCTAATCTCTCAATCTCTGACCTTTTCTC 80152
27 ..... ValLeuSerGlyHisSerLeuGlnThrProThrGluGluGly 41
80153 TATATAGTTGTGTCGGACACTCTCTTCAGACACCTACAGAGAGGCC 80202
41 InGlyProGluGlyValTrpGlyProTrpValGlnTrpAlaSerCysSer 57
|||||

```



80203 AGGGCCCCGAAAGGTGCTCGGGACCTTGGGTCCAGTGGGCCCTCTTGCTCC 80252  
58 GlnProCysGlyValGlyValGlnArgArgSerArgThrCysGlnLeuPr 74  
80253 CAGCCCTGGGGGTGGGTGTCAGCGCAGGACCGGACATGTCACTCC 80302  
74 oThrValGlnLeuHisProSerLeuProLeuProProArgProProArgH 91  
80303 TACAGTGCAGCTCCACCCAGGTCTGCCCTTCCTCCCGGCCGCCCAAGAC 80352  
91 isProGluAlaLeuLeuProArgGlyGlnGlnGlyProArgProGlnThrSer 107  
80353 ATCCAGAGCCCTCTCCCGCGGGCCAGGTCCTCCAGACCCAGACTTCT 80402  
108 ProGluThrLeuProLeuTyrArgThrGlnSerArgGlyArgGlyGlyPr 124  
80403 CCAGAAACCTCCCTTGTACAGGACACAGTCTCGGGGAAGGGGTGGCCC 80452  
124 oLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluLea 141  
80453 ACTTCGAGGTCCCGCTTCCACCTTAGGGAGAGAGAGACCCAGGAGATTC 80502  
141 rgAlaAlaArg..... 144  
80503 GAGCGCCAGGAGGTGAGAGCCTGGGTGGAGAGGTGGCCTTGGGCAA 80552  
144 ..... 144  
80553 GGTGGGGCTTAGCATGAAGAAAGGGAGCACGGGTGGCTTCAGAGGC 80602  
144 ..... 144  
80603 TTGTCTTTTGGTGCCAGGAAGTAACTGGGGAACAGGGTCAGAGGTGGTC 80652  
145 ..... Arg\_SerArgLeuArgAs 150  
80653 AGAGCTGTGCCCCACATCTGAACACCTTCTCAGGTCCCGGCTTCGAGA 80702  
150 pProileIysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuP 167  
80703 CCCCATCAAGCCAGGAATGTTCGGTTATGGAGAGTGCCCTTTTGCAATTGC 80752  
167 roLeuHisArgAsnArgHisProArgSerProProArgSerGluLeu 183  
80753 CACTGCACCGGAACCGGAGCACCTCGGAGCCCCACCATCTGAGCTG 80802  
184 SerLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProAr 200  
80803 TCCCTGATCTCTTAGAGGGAGAGGCTATTCGTCCTTACTCCCAAG 80852  
200 gAlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProT 217  
80853 AGCAGAGCCATTCTCCGAAACGGCAGCCCCCAAACTGAGTCCCTCCCA 80902  
217 hrGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerPro 233  
80903 CAGAACTGTCTGTCCACACCCCTTCCCGCCCAAGCAGAACCTCTAAGCCCT 80952  
234 GluThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuAr 250  
80953 GAAACTGTCTCAGACAGAGGTGGCCCCCAGAACACAGGCTGCCCCCTACG 81002  
250 gHisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrH 267  
81003 GCATCACCCAGAGCCAGGCTCTGGGCACAGAGCCCCCTCACCCACAGC 81052  
267 isSerLeuGlyGlyGlyPhePheArgAlaSerProGlnProArgArg 283  
81053 ACTCCTTAGGAGAGGTGGCTTCTCCGTGCATCCCTTCAGCCACGAAAG 81102  
284 ProSerSerGlnGlyThrAlaSerProGlnValAlaGlyArgArgProAs 300  
81103 CCAAGTTCACAGGTGGGTCAGTCCCGAGTACAGGGAGACGCCCTGA 81152

300 pProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProT 317  
81153 TCCTTTTCTTCGTTCCTCGGGCCGAGGCCAGCAGGGGCCAAGGCCCTT 81202  
317 rpGlyThrGlyThrProHisGlyProArgLeuGluProAspProGln 333  
81203 GGGNACGGGGGGGACTCCTCACGGGCCCGCTGAGCCTGAGCCTCAG 81252  
334 HisProGlyAlaTtpLeuProLeuLeuSerAsnGlyProHisAlaSerSe 350  
81253 CACCGGGCGCTGCTGCCCTGTAGCAACGGCCCCCATGCCAGCTC 81302  
350 rLeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyG 367  
81303 CCTCTGGAGCCTCTTGTCTCCAGTAGCCCTATTCCAAGATGTTCTGGG 81352  
367 luserGluGluLeuArgAlaCysSerGln..Ala..... 377  
81353 AGAGTGAACAGCTAAGAGCCTGCAGCCAAAGCGGTGAGTCTCCTCGGCC 81402  
377 ..... 377  
81403 CCCCTCCCAACCCCGACCTCCAGTGTGGCTTCCCTGCCCTGAAGCTGGT 81452  
377 ..... 377  
81453 CTTAGCTTCCGCTTGGCACCTGAGGAGAGGGCCTTGGCATCTGACCA 81502  
377 ..... 377  
81503 COTCAGGCGAGGGTCTTGGAGCTTTCGGTGACCTGAGCCTGCCACCT 81552  
377 ..... 377  
81553 GCTCCCGAGTTTCAGCCCTGCCCTACCTCATTTTGTCTCCCGACGCTG 81602  
378 ..... ProCysProProGluGlnProAspPr 386  
81603 ACTCTTTGTACCCCTCACCGCAGCCCTGCCCTGAGCAGCCAGACCC 81652  
386 oArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGlnL 403  
81653 CCGGGCCCTGCAGTGGCAGCCTTAACTCCAGGAATTATGCGCCAGC 81702  
403 eutyrglnTrpGluProPheThrGlu.Val..... 412  
81703 TGTATCAGTGGAGCCCTTCACTGAAGGTGAGGTTCCTTGCTCACCCCTG 81752  
412 ..... 412  
81753 GGCAGTGGTGGCTTGAATTGGGATGAAGGGAGAGAGATACCTGCTT 81802  
412 ..... 412  
81803 ACTCCAGCCCTGAATGACTTCCAGCCCTCTGCTTCCCTGGCCATGC 81852  
412 ..... 412  
81853 CTTCTTTCTTCCTCGGGCTGGGTGAGGACACAGCAGGTGGTGA 81902  
412 ..... 412  
81903 GGTCTGCAGAAAGGTGCGGGTGGGAGGAGGTGTGGAGACACGCTTT 81952  
412 ..... 412  
81953 GTCGGACTCCCTGGGAAGGCCATCACTGGGGGTGAGGTGGCTGTGACA 82002  
412 ..... 412  
82003 CAAGAGGGCCATGTTGGAGAGATCTGTCCAGCCGTGACAGCCAGGTGG 82052

412 ..... 412  
 82053 GGTTGTGCCAGCATCTCTGGGCACGTGCTGATATCGGTGGTTCCTCCAGGTTAC 82102  
 412 ..... 412  
 82103 CATCCGCTTCATTTTAGCGCTGTGTTAACTTGACACCGGGATAAGCACATC 82152  
 412 ..... 412  
 82153 AAGAGCTCTGAACACACACAGAATCATGGGCATCAGGGAATGCTCTCTCTC 82202  
 412 ..... 412  
 82203 AGTTCAGATCCCTTGATGGTCTCGGTGACCTCACTGACCTGACCACC 82252  
 412 ..... 412  
 82253 TTTTTTTTTTTTCTGAGACAGAGTCTCACTCTGTCTACCCAGGCTGGA 82302  
 412 ..... 412  
 82303 GTACAGTGGCAGCATCTCAGCTCACTGCAACCTCTGCTTCCCTGGGTTCAA 82352  
 412 ..... 412  
 82353 GCAATCTCTGCTGCTCAGCTCCCGAGTAGTGGGACTACAGGCATGTC 82402  
 412 ..... 412  
 82403 CACCACCCCTGGCTAATTTTGTATTGTAGTGGAGATGGGTTTCACCA 82452  
 412 ..... 412  
 82453 CGTTGGCCAGGCTGCTCTTGAACCTCTGACCGAATGATGCACCCACTCG 82502  
 412 ..... 412  
 82503 GCCTCCAGAGTGTGGGATTACAGGCATAGCCACTGCGCCCGGCCCTG 82552  
 412 ..... 412  
 82553 ACCACCTCAGCTTGTGCTGTGTTGCACATTTCTTCATCCACAGTGACCTG 82602  
 412 ..... 412  
 82603 GGCAACCTCAAGTGCCTCTGGGGCCCTCTCTCAGCTAACCTCCACAC 82652  
 413 ..... GlnGlySerG 416  
 82653 AGGAGGCCACTAACCCCTTCTACCCCTGTCCCTTAGTCCAGGGCTCCC 82702  
 416 lnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTyrValArg 432  
 82703 AGCGCTGTGAACCTGAACCTGCGCGCCCGCTGGCTTCGCTTATGTCCGT 82752  
 433 HisThrGluValGlnAspGlyThrLeuCysGlnProGlyAlaProAs 449  
 82753 CACACTGAAAGGTCCAGGATGGGACCCCTGTGTGAGCCTGGAGCCCTGA 82802  
 449 pileCysValAlaGlyArgCysLeu..Ser 458  
 82803 CATCTGTGTGCTGGACGCTCTCTGGTGGGGAAGACAGAGTGTGTGTGC 82852  
 458 ..... 458  
 82853 ACACACATGCATATGCACACACACATGCCCCCATATGCATACACAT 82902  
 458 ..... 458  
 82903 GTACACATATGTATGAACATGCACACATGCAAGACACATACACACAC 82952  
 458 ..... 458

82953 GCATATGGCAGACACACATGCACACACGCACACACACACACACACACAC 83002  
 458 ..... 458  
 83003 ACGCATATGCACACACATGCACATGCATATGCACACACATGCATGCACAT 83052  
 458 ..... 458  
 83053 GTGCACACGTGCATGAACACATGCACACACACACACACATGCAGACACATGCAC 83102  
 458 ..... 458  
 83103 ACACATGTAGACACACATGCATGAACACATGCACACATGCACACACACATGC 83152  
 458 ..... 458  
 83153 ATTTGCACAGTCTCAATTCAACACAAACACACATACTGAGATCACTGAG 83202  
 458 ..... 458  
 83203 GGAGAATTCTCACCTGCCAAGCTGCGCTGAGCAGAGCCTGAAGGATCTGA 83252  
 458 ..... 458  
 83253 TCGGGCACCTGTCCATGTCCCTGGGTCTGGTGGGATGGTGGGCTGTT 83302  
 458 ..... 458  
 83303 TTTGTGCTCTCACTTGTGGCACAAGCAGGGTAGTGAGCTGAGGCTCC 83352  
 458 ..... 458  
 83353 CGAGGGACCGGGGTGGGTGAGGTGTTCTGGCTTCTGGCTTCTGTGGCCACT 83402  
 459 ..... ProGlyCysAspGlyIleLeuGlyse 467  
 83403 GCCTCACCTCACTCTCCAGAGCCCGGCTGTGATGGGATCCTTGGCTC 83452  
 467 rGlyArgArgProAspGlyCysGlyValCysGlyGlyAspSerThrC 484  
 83453 TGGCAGGCTCCTGTGATGGGTGTGGAGTCTGTGGGGGTGATGATCTACCT 83502  
 484 ysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyr 500  
 83503 GTCGCTTCTTTTCGGGAACCTCACTGACCGAGGGGCCCTGGCTAT 83552  
 501 GlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaCl 517  
 83553 CAGAAGATCTTGTGGATTCCAGCGGAGCCTTGGCGCTCCAGATTGCCCA 83602  
 517 nLeuArgProSerSerAsnTyrLeu..... 525  
 83603 GCTCCGGCTAGCTCCAACTACCTGGGTGAGCACCCAGCTGCCCTCCCTT 83652  
 525 ..... 525  
 83653 CCACITTCGCTCTGTGTCGGCCCTCCATACCCCTACTCAGACAGTGAAC 83702  
 525 ..... 525  
 83703 AACCCAAACAGGGGGGAAGTCCAGGGGCTAGCCCTCCCTCGTGGNAAGGA 83752  
 525 ..... 525  
 83753 GTGAGGAAGCTGAGAGGCTTGGGGGATCTTAGTTCCTGTGGGAGCTT 83802  
 525 ..... 525  
 83803 CTATAGGCTAAGCACACGGTGTGGGAGGAGGATATTATCACCTGGA 83852  
 526 ..... Ala..LeuArgGlyProGlyGly 532  
 |||||

83853 ATTTCGCCATCTCTCAGCTCTGACCCGACGACACTTCGTGGCCCTGGGGC 83902  
 533 ArgSerIleIleAsnGlyAsnTrpAlaValaspProGlySerTyrAr 549  
 83903 CGGTCCATCATCAATGGGAACCTGGGTGGATCCCTCGGTCTCTACAG 83952  
 549 gAlaGlyGlyThrValPheArgTyrAsnArgProProArgGluGluGlyL 566  
 83953 GGCGGGGACCGGCTCTTCGATATAACCGTCTCTCCAGGGAGGAGGCA 84002  
 566 yGlyGlySerLeuSerAlaGluGlyProThrThrGlnProValaspVal 582  
 84003 AAGGGAGAGTCTGTGGGTGAAGGCCCCACCACCGACCTGTGGATGC 84052  
 583 Tyr..... 583  
 84053 TATGTGAGCCTGGGGCCAGGGCAGCTGAATGCTGGGAGGAGGCTGTT 84102  
 583 ..... 583  
 84103 CCCTCTGGCAGAGCTGTGGTTGTCAAGATGGGAGAGAGCTGGTGC 84152  
 584 .....Me 584  
 84153 TCCTCTTCTGGCCACCCCCACATATTCATATCTCTCTCTCTCTCTCT 84202  
 584 tIlePheGlnGluAsnProGlyValPheTyrGlnTyrValIleSers 601  
 84203 GATCTTTCAGGAGAAACCCAGCGCTTTTATCAGATGATCATCTCT 84252  
 601 erProProIleLeuGluAsnProThrProGluProValProGln 617  
 84253 CACCTCTCCATCTTGTGAACCCACCCAGAGCCCTGTCTCCCCAG 84302  
 618 LeuGlnPro..... 620  
 84303 CTTGAGCCGGGTAAAGCTCTGACCCCTGACCTGGAAGGAGGAGGAG 84352  
 620 ..... 620  
 84353 GCTGAGGGTGGCTGGGGCACTGGGGTGGCATCTGATTCGCTGCTCC 84402  
 621 .....GluIleLeuArgValGluProProLeuAlaProAlaProArg 634  
 84403 CCTGCACAGAGATCTGAGGGTGGAGCCCACTGCTCGGCAACCCCG 84452  
 635 ProAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGlnMe 651  
 84453 CCAGCCGGACCCAGCACCCTCCAGCGTCAAGTGGGATCCCCCAGAT 84502  
 651 tProAlaProProHisProArgThrProLeuGlySerProAlaAlaTyrT 668  
 84503 GCCCGCCCGCCCATCCAGAGACACCCCTGGGGTCTCCAGCTGCGTACT 84552  
 668 rpLysArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGly... 683  
 84553 GGAACAGAGTGGGACACTCTGCAATGCTCAGCGTCTCGCGGAAAGGTGAG 84602  
 683 ..... 683  
 84603 ACATCAGAGTGGTTCGCCCATCTCGGTCCAAACCCCCCACTGACACT 84652  
 683 ..... 683  
 84653 CCGCATCTCTGATTTGTGGGCCACGCCGACCCCTGCGTCTGGGACACCA 84702  
 683 ..... 683  
 84703 CTGAGCTTGGGCTCTAGCTCTGCTCCCTGCTGCTGCTCTCTCTCTCTCT 84752  
 684 ...Val...TrpArgProIlePheLeuCysIleSerArgGluSerGlyGlu 698  
 84753 CAGGTGCTGCGCGCCCATTTCTCTCTGCACTCTCCCGTGTGAGTGGGAGAG 84802

699 GluLeuAspGluArgSerCysAlaAlaGlyValaArgProProAlaSerPr 715  
 84803 GAACCTGGATGAACCCAGCTGTGCCGCGGTGCCAGGCCCCAGCTCCCC 84852  
 715 oGluProCysHisGlyThrProCysProPro..... 725  
 84853 TGAACCTGGCAGCGCACCCCATGCCCCCATAGTGTAGTATGGGGGAGCC 84902  
 725 ..... 725  
 84903 CACGGGAGGGTTAGGTACTGTGAACACACAGCAGTTGCCCCAGAACTTT 84952  
 725 ..... 725  
 84953 ACCTGAACCTCTGCATCAACGCAACATCTCTCTGCTGCAGAGAGACA 85002  
 725 ..... 725  
 85003 TTAGATTAGAAGGCAGACATTCGGTAGGCAGCAAGAAAGCATGGCAGG 85052  
 725 ..... 725  
 85053 CTGAGCCCCCAGGAACCCAGACTCCAAACGCCAAACGTGCCCACTCTCC 85102  
 725 ..... 725  
 85103 TCTGAGGCCCCAGCTCCAGTGTCTTCTTCCATCTCTGCTGCCACGGCAGC 85152  
 725 ..... 725  
 85153 CCCACACATCTCATATGTCTCGCCCTCTTCCCTGCGCTGCTGCTGCT 85202  
 726 ...TyrTrpGluAlaGlyGlnTrpThrSerCysSerArgSerCysGlyPro 741  
 85203 CGAGCTGGAGGCTGGCAGTGGAGATCTCTGACCCGCTCTCTGTGGCCCC 85252  
 742 GlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGlyGly 758  
 85253 GGCACCCAGCACCCGACGCTGCAGTGCAGGAGAAATTTGGGGGGGTGG 85302  
 758 ySerSerValProProGluArgCysGlyHisLeuProArgProAsnIleT 775  
 85303 CTCTCTGGTGGCCCGAGCGCTGTGGACATCTCCCCCGCCCAACATCA 85352  
 775 hrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluValGlySerPro 791  
 85353 CCCAGTCTGCCAGCTGCGCTCTGTGCCCATTTGGGAAGTTGGCTCTCT 85402  
 792 Trp..... 792  
 85403 TGGAGCCAGGTGAGTTTCCCCAGGCAAGAGGTGCTGGGGAGGGAATGG 85452  
 792 ..... 792  
 85453 GCACAGGTGAACACAGCAGTGGTTTTTCAACTCTCTTTTTCATCAGAGA 85502  
 792 ..... 792  
 85503 AACTATTTACAAACAGGGTTTCATATGGACCCCAATATAGAAGTCAGAT 85552  
 792 ..... 792  
 85553 AAGGAGCTGAACCCAGGACTGGGGGCCCGCCAGGCCCTAGGGCTCAGCTCC 85602  
 792 ..... 792  
 85603 TCCTCAGCCCCAGAAAGCCATGTGACTGTGGGTGGGGATCGAAGGCAGA 85652  
 792 ..... 792  
 85653 GCGTGTCTTGAAGCCTAGAGCTGGAAGCTGCTGAGAGGTCTGGGAG 85702

793 .....Ser..GlnCysSerValArgCysGly 800  
85703 CCCAGCTCCCTGGAATCCCTCGCCCTCAGTCTCGGTGGTGGCC 85752  
801 ArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspG1 817  
85753 CGGGCCAGAGAGAGCGGAGGTCCTGCTGTGGGAAACATGGTGATGA 85802  
817 uValSerGluGlnGluCysAlaSerGlyProGlnProGlnProSerArgG 834  
85803 ACTGAGCAGCAGGAGTGGCTGAGGCCGCCGCCGCCGCCGCCGCCAG 85852  
834 luAlaCysAspMetGlyProCysThrAlaTrpPheHisSerAspTrp 850  
85853 AGGCCTGTGACATGGGGCCTGTACTGCTGTGGTTCACAGCCAGCTGG 85902  
851 SerSerLysValSerProGluProAlaIleSerCysIleLeuGlyAs 867  
85903 AGCTCCAGGTGAGCCGGAACCCCGCCGACCATATCTGCTGCTGGTAA 85952  
867 nHisAlaGlnAspThrSerAlaPheProAla 877  
85953 CCAGGCCAGGACGCTCAGCCTTTCCAGCA 85983  
seq\_name: gb\_htg:AC092479

seq\_documentation\_block:  
LOCUS AC092479 142023 bp DNA linear HTG 05-FEB-2002  
DEFINITION Mus musculus clone rp23-231115 strain C57BL/6J, WORKING DRAFT  
SEQUENCE, 3 unordered pieces.  
ACCESSION AC092479  
VERSION AC092479.12 GI:18497125  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Jia.H., Zhang.P., Lin.S., Wu.H. and Roe.B.A.  
Mus musculus BAC Clone rp23-231115  
Unpublished  
2 (bases 1 to 142023)  
Jia.H., Zhang.P., Lin.S., Wu.H. and Roe.B.A.  
Direct Submission  
Submitted (10-JUL-2001) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Feb 5, 2002 this sequence version replaced gi:18390246.  
----- Genome Center  
Center: Department of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 32949: contig of 32949 bp in length  
\* 33049: gap of unknown length  
\* 33050 78918: contig of 45869 bp in length  
\* 78919 79018: gap of unknown length  
\* 79019 142023: contig of 63005 bp in length.  
Location/Qualifiers  
1. 142023  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="rp23-231115"  
/clone\_lib="RPCI mouse BAC library 23"

BASE COUNT 37523 a 34071 c 31510 g 38713 t 206 others  
ORIGIN  
alignment\_scores:  
Quality: 2670.50 Length: 1750  
Ratio: 3.826 Gaps: 28  
Percent Similarity: 39.886 Percent Identity: 35.371  
alignment\_block:  
US-10-041-770-2 x AC092479 ..  
Align seg 1/1 to: AC092479 from: 1 to: 142023  
28 LeuSerGlyHisSerLeuGlnThrProThrGluGluGlnGlyProG1 44  
81735 CTGTTTGGACCTTCTCATCAGCTCCATCAGAGGAGGCGCCCTGA 81784  
44 uGlyValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysG 61  
81785 AGGCCTCTGGGGACCTTGGGTGCGATGGCTTCTGCTCCAGCCCTGTG 81834  
61 lYValGlyValGlnArgSerArgThrCysGlnLeuProThrValGln 77  
81835 GGGTGGGGTGCAGCGAGGAGCGCAATGT.....GAA 81869  
78 LeuHisProSerLeuProLeuProProArgProArgHisProGluA1 94  
81870 CTGACCCAGCCCTACCCCTCCCTCCCGAGCCCAAGACATCCAGAAGC 81919  
94 aLeuLeuProArgGlyGlnGlyProArgProGlnThrSerProGluThrL 111  
81920 CCACGGGCGGGGGCCAGGCTCCAGACCCAGCTCCAGGGATCCCC 81969  
111 euProLeuTrpArgThrGlnSerArgGlyArgGlyGlyProLeuArgGly 127  
81970 AGTCTCTGTATAGGCCACAGCCTCGGGAGAGGGTGGCCCTCTTCGAGCT 82019  
128 ProAlaSerHisLeuGlyArgGluGluThrGlnGlu.IleArgAlaAla 144  
82020 CCGTCTTCTCAAGTGGGAGAGAGAAACCCAGGAGCTCAAGGGGGCCA 82069  
144 T9..... 144  
82070 GAGGTGAGGCTCTGAGGCAGAGCAGACAGGGGGTAAATGGAAGGCCCT 82119  
144 ..... 144  
82120 TCAGCAGGCCAATCTTTCAGGAAGCAGACTACGGAGACAGGCTAAGAGAT 82169  
145 .....ArgSerArgLeuArgAspProIleLysP 154  
82170 GACTAAGTCTACATCTCTCTAGGTTCGGGTTCGAGACCCCATCAAGC 82219  
154 roGlyMetPheGlyTyrglyArgValPropheAlaLeuProLeuHisArg 170  
82220 CAGGGATGTCGTTATGGAGAGTGCCTTTTGCCTTGCTCGCACCGG 82269  
171 AsnArgArgHisProArgSerPro.....ProArgSerGluLeuSerle 185  
82270 AGCCCGAGGCATCTCTACACAGCCCGGGCAACCCCAAGAAC..... 82308  
185 uIleSerSerArgGlyGluGluAlaIleProSerProThrProArgAlaG 202  
82309 ....TCTTCTACAGGGGAGGGATGTTCCATCTCAGCCTCCCAAGCAGAG 82354  
202 luProPheSerAlaSerGlySerProGlnThrGluLeuPro...ProThr 217  
82355 AGCTGGCCTCTGAAAAGACAGCGCCCTCAGATGACGCCCTGAAACCCCGT 82404  
218 GluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG1 234  
82405 TCTACTAGCTGAAACTCCA.....AGGCTCGG 82433

412	.....	412
83265	GGAGGGATTTTGGAACTGATGGGTGGATGGTCAAGACTGGAGCAGGCT	83314
412	.....	412
83315	GGGAGGGAGCTGTCTAGAGAGCTCTTTCCGACAGGGCGCTGACTGT	83364
412	.....	412
83365	AACCTGTTCCAGCAGGACACTGTTGCTCCCAAGTTACTGTCCACTGGAC	83414
412	.....	412
83415	TTGTGGCTATGTCAACTCCACACCAAGACAAACCTTCATGAGCTATGAAC	83464
412	.....	412
83465	TCATGATCACAGTGTCCAGAAACCCCTTTTGAGGAATGGTAGGCTTGCTC	83514
412	.....	412
83515	AGTGGGTAAAGGAAGCTTGTCTCAAGCCTAGTGACCTGAGTTTGAICTCCA	83564
412	.....	412
83565	GAATCCAGAACCGCTCCTGCAGATGACCTCTTCTCCCACTGAGCA	83614
412	.....	412
83615	CCCCAATCTTATACACAAAGGAAAAACAAACAAAAATCTTTGATGCC	83664
412	.....	412
83665	AGCTGCCAAACTTTCCTTTTTTAGGAGGACGAATGGAAGAACAGTCA	83714
412	.....	412
83715	AATCTCTTTTHTTTTTTTTTTTTTTTTTTTTTTCTGAAACGGGGTTTC	83764
412	.....	412
83765	TCTGTGCAGTCTGGCTGTCTCTGGAACCTCACTCTGTAAACAGAGCTGGCC	83814
412	.....	412
83815	TCAAACTCAAAAATCTGCTTGCCTCTGCCTCCCAATGCTGGGATTAAG	83864
412	.....	412
83865	GTGTGCACCAACCAGCTGGCATCAGTCAAAATCTTAACTCGAGACCAC	83914
412	.....	412
83915	ATAGCATACAAAACATAAGACACAAAATCGTAAGCGTGGGAAGTTGC	83964
412	.....	412
83965	CCCTCCCCCCTTCCACAGCAGCTTCCGATGCCCCCATATGTAAAGGCT	84014
412	.....	412
84015	GTCAAGGCTCTTGGGGGGGTGGCTCTTTCTCGAGGGAACCCAGTTAAC	84064
413	.....GlnGlySerGlnArgCysGluLeuAs 	421
84065	TAGCTGTGTCCTCTCCCTTTAGTTCAGGGCTCCCAAGCGCTCTGAACCTGA	84114
421	nCysArgProArgGlyPheArgPheTyrValArgHisThrGlnLysValG 	438
84115	CTGGCGCCCCGTGGCTTCCGATTCTATGTCGACACACTGAAAGGGTTC	84164
438	lnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysValAlaGly	454

|||||  
84165 AGGATGAACCTGTGTACGCTGGATCCTAGACATCTGTGTGGCTGA 84214  
455 ArgCysLeu..Ser..... 458  
84215 CGTGCTGTGTAGGAGGAGGACGGGTGTGGTTGTGTACATGTGCACCTAC 84264  
458 ..... 458  
84265 ACACACACACACACACACACACACACACACACACACCTCTTTAATTGC 84314  
458 ..... 458  
84315 CACAAGAGTGAAAAACACACAGACTGAGGTGGCGAAAGGAAAGTTTCTGCT 84364  
458 ..... 458  
84365 GTCCAGTAAGGATGGGTGTAGACAGGATCCCAACAGACTTGGCTAAAGC 84414  
458 ..... 458  
84415 TCCTCCAGACATAGTGGGGAGGAGGAGGTCAGAGAATACTCTGACTTGA 84464  
458 ..... 458  
84465 AGCACAAAGAGTGTGCAGCAGGAGGTGAGCTGAGGGCCCTGGGTGGAGC 84514  
458 ..... 458  
84515 GGGGCCAGGTGTGAGGATGCCGTGGCTGTGTGGCCCTGCCTCACCT 84564  
459 .....ProGlyCysAspGlyTleLeuGlySerGlyArgar 470  
84565 CACTCTCCCAAGAGCCCTGGCTGTGTAGTGGGTCTTGGCTGGCAGGCG 84614  
470 qProAspGlyCysGlyValCysGlyGlyAspAspSerThrCysArgLeuV 487  
84615 TCCGATGGTGGGAGTCTGCAGGGGTGATGTTTACTCTGCCGCTGG 84664  
487 alserGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLysIle 503  
84665 TTTCGGAAATCTCACTACCGAGGGGGCCCTTGGCTATCAGAAATC 84714  
504 LeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgpr 520  
84715 CTGTGGATCCCTGCCGGGGCTCCACCTTCACATTTCCAGTTGGACC 84764  
520 oSerSerAsnTyrLeu..... 525  
84765 CAGTTCCAATTACTCGGTAAAGAAAGAGGACCTCGGTCTCTATGCCCCAGT 84814  
525 ..... 525  
84815 CTATCCTTTTCTTTGATCTGCCTACTACACACCTCTTCAGCCGAACGA 84864  
525 ..... 525  
84865 AAGCNAACAGGAGAGGTGGAGCCTTCTCATAAACCGATGGAGGT 84914  
525 ..... 525  
84915 GAGGAGGTGGATGTCAGGTCTCTGGGCACAAAGGCTTCAGAAAGAGAG 84964  
526 .....AlaLeuArgGlyPro 530  
84965 TACCACATCACTCAGCTCAGACATCTGTCTTTCAGCAGCTCCGAGGCT 85014  
531 GlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspProGlyse 547  
85015 GGGGGCGCTCCATTATCAATGGGAAGTGGGCTGTGGATCCTCCAGGCTC 85064  
547 rTyrArgAlaGlyGlyThrValPheArgTyrAsnArgProProArgGluG 564  
||||| ||| |||||

85065 CTATACAGCCATTGGGACTGTCTTCCAGTATAACCGTCTCCACGGGAAG 85114  
564 luGlyLysGlyCylUserLeuSerAlaGluGlyProThrThrGlnProVal 580  
85115 AAGCAAGGGGAGAGTCTGTACAGAAAGCCCTACCACCTCAGCCTGTG 85164  
581 AspVal..Tyr..... 583  
85165 GACGCTCTATGTGAGCTGGGCGAGGTGTAGATCGTTAGGGAGGGAGGAG 85214  
583 ..... 583  
85215 GGCCCTTAGGGCAGACAGTGTGGATCTAGCCTAGCCTATCCCTCTCTG 85264  
584 .....MetIlePheGlnGluG1 589  
85265 TTACTCCAAAGCCATTGTGATCTTCCCTCCTCAGATGATCTTTCAGAGGA 85314  
589 uAsnProGlyValPheTyrGlnTyrValIleSerSerProProIleL 606  
85315 CAACCCAGGTGTCTTTTATCAGTATGTCATCTCTCCCTCCTGCAGTCC 85364  
606 euGluAsnProThrProGluProProValProGlnLeuGlnPro..... 620  
85365 TAGAGAGTCTTCCACAAAGCCTCCAGCCCTCAGCCTCAGCCTCGGTAAA 85414  
620 ..... 620  
85415 ACTAGGATCCCGACAGCCAGGAGAGGGGTGAGCGAGGACTGAGCTCC 85464  
621 .....Glu..11 622  
85465 AGGGTAGCAGGAGGAGCATCTAATCTGGTCCCTGGATTCCCTCGCAGAGAT 85514  
622 eleuArgValGluProProLeuAlaProAlaProAlaArgThrP 639  
85515 GCTGAGGGGAGGCCCTTACTCCCTCAGCCCGCCAGTCCGGGCAC 85564  
639 roGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProPro 655  
85565 CAGGCACCTCCACGTCAGGTGCGGATCCCAAGTCCCTCCCTCCAGT 85614  
656 HisProArgThrProLeuGlySerProAlaAlaTyrTrpLysArgValG1 672  
85615 CCGTCAGGACAGCCATGGGTCTTCTGTGGATACTGGAACAAGTGGG 85664  
672 yHisSerAlaCysSerAlaSerCysGlyLys..... 682  
85665 GCACCTCTCAGTGTTCAGCATCCTGTGGCAAGGTGAAGGAGACACCTCA 85714  
682 ..... 682  
85715 GCCCAATCCCTGTTCTCTATATATATCTATACAGATTCTGTATCCTGGA 85764  
682 ..... 682  
85765 CCACGGGTCTACATTGGACCTATGCTGGGATTTTCATTGAACGGGCT 85814  
683 .....GlyValT 685  
85815 CTGAGCCTTCCTGCTCCCTCCCTGCTGCTCTTATTCCTTCAGGTGTTT 85864  
685 rpArgProIlePheLeuCysIleSerArgGluSerGlyGluLeuLeuasp 701  
85865 GGCACCCCAATTTCTCTGCAATTCCTGCTGAGTCAGAGAGAGTGGAT 85914  
702 GluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluProCy 718  
85915 GAACAGAGTGTGTGGGGCCAGACCCCGAGCTTCCCTGGAACCTGTG 85964  
718 shIsGlyThrProCysProPro..... 725  
85965 CCATGSAACCCCGTGTCTCCATAGTGTAGTACGGGCCACAGGAGGGCTG 86014  
||||| |||||







111495 CTTGTGGCTATGTCACCTCCACACCAAGACAAACCTTCATGAGCTATGAA 111544  
412 ..... 412  
111545 CTCATGATCACAGTGTGAGAAACCTTTTGAGGAATGAGCTGGCT 111594  
412 ..... 412  
111595 CAGTGGGTAAAGAAAGTGTCTCAAGCCTAGTGACCTGAGTTGATCTCC 111644  
412 ..... 412  
111645 AGAATCCAGAACCCGCTCTCTCAAGATGACCTCTTCTCCACCTGAGC 111694  
412 ..... 412  
111695 ACCCAATCTTATACAAAGGAAACAAACAAACAAATCTTTGATGSC 111744  
412 ..... 412  
111745 CAGCTGCCAACTTCCCTTTTITAGGAGGACGAATGGAAGAACAGTC 111794  
412 ..... 412  
111795 AAATCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTGAGACGGGTTT 111844  
412 ..... 412  
111845 CTCTGTGCAGTCTCTGCTGCTGCGAACCTCTCTGACACAGCTGGC 111894  
412 ..... 412  
111895 CTCAAACTCAGAAATCTGCTGCTGCTGCCCTCCAAATGCTGGATTAA 111944  
412 ..... 412  
111945 GGTGTGCACCAACCGCTGGCATCATCTTAACCTTAACCTGACACCA 111994  
412 ..... 412  
111995 CATAGCATAACAAAAACATAAGACACAAATCGTAAGCGTGGAAAGTTG 112044  
412 ..... 412  
112045 CCGCTCCCGCTTCCACACAGCTTCCCGATGCCCGCATATGTAAGGC 112094  
412 ..... 412  
112095 TGTCAGGCTCTTGGGGGGGGTGGCTTTCTCGAGGGAACCCAGTTAA 112144  
413 ..... GlnGlySerGlnArgCysGluLeuA 421  
112145 CTAGCCTGTGTCTCTCTCTTTAGTTACAGGCTCCACGCGTGTGAAGTGA 112194  
421 snCysArgProArgGlyPheArgPheTyrValArgHisThrGluLysVal 437  
112195 ACTCGCGCGCGTGGTTCGGATCTCATGTCGACACACTGAAAGGTG 112244  
438 GlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysValAlaG1 454  
112245 CAGGATGGAACCTGTGTACGCTGGATCCCTAGACATCTGTGTGGCTGG 112294  
454 YArgCysLeu...Ser..... 458  
112295 ACGTGTGCTGAGGAGGCGAGGCTGTGGTTGTATGATGTGCACCTA 112344  
458 ..... 458  
112345 CACACACACACACACACACACACACACACACACACTCCCTTAATTTG 112394  
458 ..... 458  
112395 CCACAAGAGTGAAACACAGACTGAGGTGGCCAAAGGAAAGTTTCTGC 112444

458 ..... 458  
112445 TGTCAGTAAAGATGGTGTAGACAGATCCACAGACTTGGCTAAAG 112494  
458 ..... 458  
112495 CTCCTCCAGACATAGTGGGAGGAGGAGGTTCAGAGATACTCTGACTTG 112544  
458 ..... 458  
112545 AAGCACAAGAGTGTGGCAGCAGGAGGTGAGCTGAGGCGCTGGGTGGAG 112594  
458 ..... 458  
112595 CGGGCCAGGTGTGAGGATCCCTGGCTGTGTGGCCCTGCCTCACC 112644  
459 ..... ProGlyCysAspGlyIleLeuGlySerGlyArgA 470  
112645 TCACCTCTCCACAGAGCCCTGGCTGTGATGGGTCTTGGCTCTGGCAGGC 112694  
470 rqpToAspGlyCysGlyValCysGlyGlyAspAspSerThrCysArgLeu 486  
112695 GTCCGATGGCTGGAGTCTCGGGGGTGTGTTCTACCTGCCGCTG 112744  
487 ValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLysII 503  
112745 GTTTCGGAAATCTCACTGACCGAGGGGCGCTTGGGCTATCAGAAAT 112794  
503 steuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgP 520  
112795 CTTGTGGATCCCTGCCGGGGCTCCACCTTCACATTTCCAGTTGGAC 112844  
520 toSerSerAsnTyrLeu..... 525  
112845 CCAGTTCCAATTACCTCGGTAAGAAAGAGGACCTCGGTCTCTATGCCCGAG 112894  
525 ..... 525  
112895 TGTATCTTTTCTTTGATCTGCTTACACACCTCTTACGCGGAACG 112944  
525 ..... 525  
112945 AAGCAAAACAGGAAGGCTGGAGCCTTCTCATAAACCGATGGAGGG 112994  
525 ..... 525  
112995 TGAGGAGGTGGATGTCAGGTCTCTGGGCACAAAGGCTTCAGAGAGA 113044  
526 ..... AlaLeuArgGlyPr 530  
113045 GTACCACTATCACTCAGCTCAGACATCTGCTTTTCAGCAGCTCCGAGGGCC 113094  
530 oglyGlyArgSerIleleAsnGlyAsnTrpAlaValaspProGlyS 547  
113095 TGGGGCGCTCCATTATCAATGGAACTGGGCTGTGGATCTCCAGGT 113144  
547 erTyrArgAlaGlyThrValPheArgTyrAsnArgProArgGlu 563  
113145 CCTATACGCCATTTGGAGTCTTCCAGTATACCGTCTCCACGGGA 113194  
564 GluglyLysGlySerLeuSerAlaGluGlyProThrGlnProva 580  
113195 GAAGCAAGGGGAGAGTCTGTGACAGAAAGGCCCTACCACTCAGCCTGT 113244  
580 lAspVal.Tyr..... 583  
113245 GGACGCTTATGTAGCTGGGCGAGGTGTAGATCGTTAGGGAGGGAGA 113294  
583 ..... 583  
113295 GGGCCCTTAGGCGACAGCAGTGTGGATCTAGCCTAGCGTATGCCCTCTT 113344



AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (16-APR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On Aug 12, 2000 this sequence version replaced gl:8439979.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0243G22  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Chemistry: Dye-terminator ET; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 192323 bases at least Q40  
Consensus quality: 196320 bases at least Q30  
Consensus quality: 198661 bases at least Q20  
Insert size: 204000; agarose-fp  
Insert size: 204215; sum-of-contigs  
Quality coverage: 6.36 in Q20 bases; agarose-fp  
Quality coverage: 6.31 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
-----

1 1381: contig of 1381 bp in length  
\* 1382 1481: gap of unknown length  
\* 1482 2602: contig of 1121 bp in length  
\* 2603 2702: gap of unknown length  
\* 2703 4013: contig of 1311 bp in length  
\* 4014 4113: gap of unknown length  
\* 4114 5331: contig of 1218 bp in length  
\* 5332 5431: gap of unknown length  
\* 5432 6622: contig of 1191 bp in length  
\* 6623 6722: gap of unknown length  
\* 6723 8166: contig of 1444 bp in length  
\* 8167 8266: gap of unknown length  
\* 8267 9674: contig of 1408 bp in length  
\* 9675 9774: gap of unknown length  
\* 9775 10943: contig of 1169 bp in length  
\* 10944 11043: gap of unknown length  
\* 11044 12198: contig of 1155 bp in length  
\* 12199 12298: gap of unknown length  
\* 12299 14193: contig of 1895 bp in length  
\* 14194 14293: gap of unknown length  
\* 14294 15717: contig of 1424 bp in length  
\* 15718 15817: gap of unknown length  
\* 15818 16935: contig of 1118 bp in length  
\* 16936 17035: gap of unknown length  
\* 17036 18375: contig of 1340 bp in length  
\* 18376 18475: gap of unknown length  
\* 18476 19862: contig of 1487 bp in length  
\* 19863 20062: gap of unknown length  
\* 20063 21169: contig of 1107 bp in length  
\* 21170 21269: gap of unknown length  
\* 21270 22452: contig of 1183 bp in length  
\* 22453 22552: gap of unknown length  
\* 22553 24582: contig of 2030 bp in length  
\* 24583 24682: gap of unknown length  
\* 24683 26383: contig of 1701 bp in length  
\* 26384 26483: gap of unknown length  
\* 26484 27710: contig of 1227 bp in length  
\* 27711 27810: gap of unknown length

27811 29195: contig of 1385 bp in length  
\* 29196 29295: gap of unknown length  
\* 29296 30800: contig of 1505 bp in length  
\* 30801 32555: contig of 1655 bp in length  
\* 32556 32855: gap of unknown length  
\* 32856 34533: contig of 1998 bp in length  
\* 34534 37122: gap of unknown length  
\* 37123 37222: gap of unknown length  
\* 37223 39506: contig of 2284 bp in length  
\* 39507 39606: gap of unknown length  
\* 39607 41320: contig of 1614 bp in length  
\* 41321 42978: contig of 1638 bp in length  
\* 42979 43078: gap of unknown length  
\* 43079 45773: contig of 2695 bp in length  
\* 45774 45873: gap of unknown length  
\* 45874 55056: contig of 9183 bp in length  
\* 55057 55156: gap of unknown length  
\* 55157 67059: contig of 11903 bp in length  
\* 67060 67159: gap of unknown length  
\* 67160 80260: contig of 13101 bp in length  
\* 80261 80360: gap of unknown length  
\* 80361 96528: contig of 16168 bp in length  
\* 96529 96628: gap of unknown length  
\* 96629 114529: contig of 17901 bp in length  
\* 114530 114629: gap of unknown length  
\* 114630 133673: contig of 19044 bp in length  
\* 133674 133773: gap of unknown length  
\* 133774 153791: contig of 20018 bp in length  
\* 153792 153891: gap of unknown length  
\* 153892 176079: contig of 22188 bp in length  
\* 176080 176179: gap of unknown length  
\* 176180 207815: contig of 31636 bp in length.  
----- Location/Qualifiers -----  
1. 207815  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-243G22"  
misc\_feature 1..1381  
/note="assembly\_name:Contig13"  
misc\_feature 1482..2602  
/note="assembly\_name:Contig15"  
misc\_feature 2703..4013  
/note="assembly\_name:Contig21"  
misc\_feature 4114..5331  
/note="assembly\_name:Contig25"  
misc\_feature 5432..6622  
/note="assembly\_name:Contig33"  
misc\_feature 6723..8166  
/note="assembly\_name:Contig37"  
misc\_feature 8267..9674  
/note="assembly\_name:Contig38"  
misc\_feature 9775..10943  
/note="assembly\_name:Contig40"  
misc\_feature 11044..12198  
/note="assembly\_name:Contig42"  
misc\_feature 12299..14193  
/note="assembly\_name:Contig45"  
misc\_feature 14294..15717  
/note="assembly\_name:Contig47"  
misc\_feature 15818..16935  
/note="assembly\_name:Contig50"  
misc\_feature 17036..18375  
/note="assembly\_name:Contig51"  
misc\_feature 18476..19962  
/note="assembly\_name:Contig52"  
misc\_feature 20063..21169  
/note="assembly\_name:Contig53"  
misc\_feature 21270..22452  
/note="assembly\_name:Contig54"

458	.....	458
133419	CCCATATGCATACACATGTACACACATATGTATGAAACACATGCACACAT	133370
458	.....	458
133369	GCAAGCACAATACACACACGCATATGCGCACAGACACATGCACACACGCAC	133320
458	.....	458
133319	ACACACACGACACACACGCATATGCACACACATGCACATGCACATATGC	133270
458	.....	458
133269	AGACACATGCATGCACATGTGCACACGTGCATGAACACATGCACACACAC	133220
458	.....	458
133219	GTATGCACACATGCACACACATGTAGACACACATGCATGAACACATGC	133170
458	.....	458
133169	ACACATGCACACACACGTATTTGCACACAGTCTCAATTCAACAACAAACAC	133120
458	.....	458
133119	CATACTCAGATCACTGAGGGAGATTCTCACTGCCAAGCTCGCGTGAGC	133070
458	.....	458
133069	AGAGCCTGAAGGATCTGATCGGGCACCTGTCATGTCCCTGGGTGTGGCT	133020
458	.....	458
133019	GGGGATGGTGGGGCTGTTTTTGTGCTCACTGTGGGCACAAAAGCAGG	132970
458	.....	458
132969	GTAGTAGCTAGAGCTCCGAGAGGACCGGGTGGGTTGAGGTGGTGTGC	132920
459	.....ProGlycy	461
132919	TGGCGTTCTGTGGCCACTGCCTCACTCACTCTCTCCAGAGCCCGGCTG	132870
461	sAPGlyIleLeuGlySerGlyArgArgProAspGlyCysGlyValCysG	478
132869	TGATGGGATCCTTGGCTCTGGCAGCGCTCTGATGGCTGTGGAGTCTGTG	132820
478	lyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArg	494
132819	GGGTGTGATTTCTACCTGTCGCTTGTTTCGGGGAACCTCACTGACCGA	132770
495	GlyClyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaIle	511
132769	GGGGGCCCTGGGCTATCAGAAATCTGTGGATTCCAGCGGGGACGGCT	132720
511	uArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeu	527
132719	CGGCTCCAGATTGCCACGCTCCGGCTAGCTCCAACCTACCTGGGTGAGC	132670
527	.....	527
132669	ACCCAGTGTGCTCCCTTCCACTTCCGCTCTGTTCGGCCCTCCATACC	132620
527	.....	527
132619	CTACTCAGACGCTGACCAAGCCAAACAGGGGAGTCCAGGCCCTAGCC	132570
528	.....ArgGly.....	529
132569	CCTCCCTCGTGGAAAGAGAGTGAAGAAAGCTGAGAGGGCTTGGGGGATCTT	132520

529 ..... 529  
 132519 AGGTTCTGGTGGAGCTTCTATAGCTAAGGACACAGGTGTGGAGGAGGA 132470  
 529 ..... 529  
 132469 AGGTATTATACCTCGGAATTTCCGATCTCTACCTCTGACCCGCGAGCA 132420  
 530 .....ProGlyArgSerIleIleAsnGlyAsnTrpAlaValasp 543  
 132419 CTTCTGGCCCTGGGCGCGTCCATCATCAATGGAACCTGGGCTGTGGAT 132370  
 544 ProProGlySerTyrArgAlaGlyValThrValPheArgTyrAsnArgPr 560  
 132369 CCCCCTGGTCTCTACAGGGCGGGGACGTC.TTTCGATATACCGTCC 132321  
 560 oProArgGluGlyLysGlyGlySerLeuSerAlaGluGlyProThr 577  
 132320 TCCAGGGAGGAGGCAAGGGAGAGTCTGTGGCTGAAGGCCCCACCA 132271  
 577 hrGlnProValaspValTyr ..... 583  
 132270 CCCAGCCTGTGGATGTCTATGTGACCTGGGGCCAGGGAGCTGAATGCT 132221  
 583 ..... 583  
 132220 GGGAGGAGGCTGTTCCTCTGGCAGAGCTGTGGTTGTCAAGATGGA 132171  
 583 ..... 583  
 132170 GAGAGAGCTGTGGCATCCTCTTCTGGCCACCCACATATCATATATCT 132121  
 584 .....MetilePheGlnGluAsnProGlyValPheTyr 595  
 132120 TCTCTTCTCCCATGATCTTTCAGGAGAAACCCAGGCGTTTTTAT 132071  
 596 GlnTyrValIleSerProProProlleLeuGluAsnProThrProGl 612  
 132070 CAGTATGTCTCTTCACTCTCTCAATCTTGAGAACCCACCCAGCA 132021  
 612 uProProValProGlnLeuPro ..... 620  
 132020 GCCCCCTGTCCCCAGCTTCAGCCGGGTAAAGACTGTGACCCCTGCACITG 131971  
 620 ..... 620  
 131970 GAAGGAGGAGGAGAGGCTGCAGGGCTGCTCGGGGCGAGTGGGGTGGCAT 131921  
 621 .....GluIleLeuArgValGluProProle 629  
 131920 CTGATTCTGGCTGCTCCCTGCACAGAGATCTGAGGGTGGAGCC.CCACT 131872  
 629 uAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnArgGlnV 646  
 131871 TGTCCGGCACCCCGCCAGCCGCGGACCCAGGACCCCTCCAGGCTCAGG 131822  
 646 aArgIleProGlnMetProAlaProProHisProArgThrProLeuGly 662  
 131821 TGGGGATCCCCAGATGCGCGCCGCCCCCATCCAGGACCCCTGGGG 131772  
 663 SerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerAlaSe 679  
 131771 TCTCCAGCTCGGTACTGGAAACAGTGGACACTCTGCATGTCTCAGCGTC 131722  
 679 rCysGlyLysGly ..... 683  
 131721 CTGGGAAAGGTGAGACATCACAGTGCCTCCCGCATCTCGGTCCAAA 131672  
 683 ..... 683  
 131671 CCCCCCAACTGACATCCCGCATCTCTGGATTGTGGGGCCAGCCCGGACCC 131622  
 683 ..... 683

131621 TGGCTCTGGGACACCACTAGCTTGGGCTCTAGCTCTCTGCTCCCTGGC 131572  
 684 .....Val..TyrArgProIlePheLeuCysIleSe 693  
 131571 TGCCTTCTCACCCACTCAGGTGTCTGGCGCCCATTTCTCTGTGCATCTC 131522  
 693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710  
 131521 CCGTAGTCGGGAGAGAACTGGATGAACGACAGTGTGCCCGGGTGCCA 131472  
 710 rGProProAlaSerProGluProCysHisGlyThrProCysProPro... 725  
 131471 GGCCCCAGGCTCCCTGTGAACCTGCGCACGGCACCCCATGCCGCCCATAG 131422  
 725 ..... 725  
 131421 TGAGTATGGGGAGCCACCGGGAGGGTTAGGGTACTTGGAAACACAGCAG 131372  
 725 ..... 725  
 131371 TTGCCCCCAGAATCTTACCTGAACCTCTGCACCTCAAGCAACATCTCTCTGG 131322  
 725 ..... 725  
 131321 CTGCAGAGAGACAACATTAGATTAGAAAGCAGACATTCGGTAGGAGCA 131272  
 725 ..... 725  
 131271 AGAAGCCATGGCAGGCTGAGCCCCCAGGACCCAGACTCCAAACGCCA 131222  
 725 ..... 725  
 131221 AACGTGCCACTCTCTCTGAGGCCGCCAGTCCAGTGTGTCTTCCATCT 131172  
 725 ..... 725  
 131171 CTGTGTCACGGGAGCCCCACACATCTCATCTATGTCTCGCCTCTTCCC 131122  
 726 .....TyrTrpGluAlaGlyGluTrpThrSerCysSe 736  
 131121 TGCCTGTGTGTGCTGCAGCTGGGAGGCTGGCGAGTGGACATCTCTGCAG 131072  
 736 rArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnG 753  
 131071 CCGTCTCTGTGGCCCCGGCACCCAGCACCCGACCTGCAGTGCAGTCCGGCAGG 131022  
 753 luPheGlyGlyGlySerSerValProProGluArgCysGlyHisLeu 769  
 131021 AATTGGGGGGTGGCTCTCTCGGTGCCCCGAGCGCTGTGGACATCTC 130972  
 770 ProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHisTr 786  
 130971 CCGCGGCCAACATCACCCAGTCTTGCCAGCTGCGCTCTGTGGCCATTG 130922  
 786 pGluValGlySerProTrp.Ser ..... 793  
 130921 GGAAGTGTGCTCTCTTGGAGCCAGGTGAGTTTGCACGAGCAAGAGGTG 130872  
 793 ..... 793  
 130871 CTGGGAGGGAATGGGCACAGGTGAACAACAGCAGTGGTTTTTCAACTT 130822  
 793 ..... 793  
 130821 CTTTTCATCAGCAGAAACTATTACAAACAGGTTTCATATGGACCCCC 130772  
 793 ..... 793  
 130771 AATATAGAAGTCAGATAAGGAGCTGAACAGGAGTGGGGGCCAGGCC 130722  
 793 ..... 793

```
130721 TAGGGTCAGCCTCCCTCCTCAGCCAGGACCATGTGCTGCTGGGCTG 130672
793 ..... 793
130671 GGGATCAGAGGCAGAGCCTGGTCTGAAGCCTAGAGCTGGAAGCTGGCTG 130622
794 .....GlnCys 795
130621 AGAAGTCTGGGAGCCAGCTCCCTGGATCCCTCGCCCTCAGTGC 130572
796 SerValArgCysGlyArgGlyGlnArgSerArgGlnValArgCysValG1 812
130571 TCGTGGGTGGCGGGCGGCGGAGAGAGCCGAGGTTCGCTGTGG 130522
812 YAsnAsnGlyAspGluValSerGluGlnGluCysAlaSerGlyProProG 829
130521 GAACAATGGTGAATGAAGTGAAGCAGGAGGTGGTGGTCCAGGCCCGCCG 130472
829 lnProProSerArgGluAlaCysAspMetGlyProCysThrThrAlaTtp 845
130471 AGCCCCCAGCAGAGAGGCTGTGAATGGGGCCCTGACTACTGCTGG 130422
846 PheHisSerAspTrpSerLysValSerProGluProProAlaIeSe 862
130421 TTCACAGCAGCTGGAGCTCCAGGTGAGCCCGAAGCCAGCCATATC 130372
862 rCysIleLeuGlyAsnHisAlaGlnAspThrSerAlaPheProAla 877
130371 CTGCATCTGGGTAAACACGCCCGCAGACCTTCAGCTTCACGA 130326
seq_name: gb_htg:AC053497
seq_documentation_block:
LOCUS AC053497 207815 bp DNA linear HTG 13-AUG-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION AC053497
VERSION AC053497.4 GI:9797901
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Waterston.R.H.
The sequence of Homo sapiens clone
Unpublished
REFERENCE 2 (bases 1 to 207815)
AUTHORS Waterston.R.H.
TITLE Direct Submission
Submitted (16-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 12, 2000 this sequence version replaced gi:8439979.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0243G22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-terminator; 100% of reads
Assembly: Dye-terminator; Big Dye; 0% of reads
Consensus quality: 192323 bases at least Q40
Consensus quality: 196320 bases at least Q30
Consensus quality: 198661 bases at least Q20
Insert size: 204000; agarose-fp
Insert size: 204215; sum-of-contigs
Quality coverage: 6.36 in Q20 bases; agarose-fp
```

```
Quality coverage: 6.31 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1381: contig of 1381 bp in length
1481: gap of unknown length
2602: contig of 1121 bp in length
2702: gap of unknown length
4013: contig of 1311 bp in length
4113: gap of unknown length
5331: contig of 1218 bp in length
5431: gap of unknown length
6222: contig of 1191 bp in length
6722: gap of unknown length
8166: contig of 1444 bp in length
8266: gap of unknown length
9674: contig of 1408 bp in length
9774: gap of unknown length
10943: contig of 1169 bp in length
11043: gap of unknown length
12198: contig of 1155 bp in length
12298: gap of unknown length
14193: contig of 1895 bp in length
14293: gap of unknown length
15717: contig of 1424 bp in length
15817: gap of unknown length
16935: contig of 1118 bp in length
17035: gap of unknown length
18475: contig of 1340 bp in length
18476: gap of unknown length
18476: contig of 1487 bp in length
20062: gap of unknown length
21169: contig of 1107 bp in length
21269: gap of unknown length
22452: contig of 1183 bp in length
22552: gap of unknown length
24582: contig of 2030 bp in length
24582: gap of unknown length
24583: contig of 1701 bp in length
24583: gap of unknown length
26483: gap of unknown length
27710: contig of 1227 bp in length
27810: gap of unknown length
29195: contig of 1385 bp in length
29295: gap of unknown length
30800: contig of 1505 bp in length
30900: gap of unknown length
32555: contig of 1655 bp in length
32556: gap of unknown length
34653: contig of 1998 bp in length
34653: gap of unknown length
34753: contig of 2369 bp in length
37122: contig of 2369 bp in length
37222: gap of unknown length
39508: contig of 2284 bp in length
39608: gap of unknown length
41220: contig of 1614 bp in length
41320: gap of unknown length
42978: contig of 1658 bp in length
43078: gap of unknown length
45773: contig of 2695 bp in length
45873: gap of unknown length
55059: contig of 9183 bp in length
55156: gap of unknown length
55157: contig of 11903 bp in length
67159: gap of unknown length
80260: contig of 13101 bp in length
80360: gap of unknown length
96528: contig of 16168 bp in length
```



```

vector_side:left"
45874. .55056
/note="assembly_name:Contig70"
55157. .67059
/note="assembly_name:Contig71"
67160. .80260
/note="assembly_name:Contig72"
80361. .96528
/note="assembly_name:Contig73"
96629. .114529
/note="assembly_name:Contig74"
114630. .133673
/note="assembly_name:Contig75"
133774. .153791

alignment_scores:
  Quality: 2014.50      Length: 834
  Ratio: 4.304          Gaps: 14
Percent Similarity: 56.115 Percent Identity: 52.998

alignment_block:
US-10-041-770-2 x AC053497 ..

Align seg 1/1 to: AC053497 from: 1 to: 207815

7 ArgProTrpLeuYrLeuLeuLeuSerLeuProGlnLeuCysLe 23
|||||
64556 AGGCCCTGGCTGATCTGCTGCTGCTTGTGTCCTCCCTCAGCTCTGCTT 64605
23 uaspGlnGlu..... 26
|||||
64606 GGATCAGGAGGTGAGTCTGGACAAGTGAGCAGCTGACGCTGCCTGCCA 64655
26 ..... 26
64656 CCCTTTCATCTCCCTTAGGTCCTCCACCCCATCTCTCCAGGCCACGCCTC 64705
26 ..... 26
64706 CATTCGCCACAGGCCACCCACCTCCCTGCCAAGTCCCGAGTTCCTGGCC 64755
26 ..... 26
64756 TCATACACCAAGAGGCCGAGGTGTAGTTCTCCCTCTGCTGCTGAATGTG 64805
26 ..... 26
64806 ACCTTGGACTTGTAGCGACTCCGTGAGCCTCAGTGTGCTCTACATGGAC 64855
26 ..... 26
64856 ACTCTGGACAGTTCCTCTCTAATCTTCAATCTCTGTCAGCCTTTTCTC 64905
27 .....ValLeuSerGlyHisSerLeuGlnThrProThrGluGluGlyG 41
|||||
64906 TATATAGGTGTGTCCGGACACTCTCTTCAGACACTCAGACAGCTACAGAGAGGGCC 64955
41 lngLysProGluGlyValTrpGlyProTrpValGlnTrpAlaSerCysSer 57
|||||
64956 AGGCCCCGAGAGGTCTGCGGACCTCTGGTCCAGTGGGCCCTCTTCTGCTCC 65005
58 GlnProCysGlyValGlyValGlnArgSerArgThrCysGlnLeuPr 74
|||||
65006 CAGCCCTCGCGGGTGGGGGTTCAGCGCAGGAGCCGACATGTCAGCTCCC 65055
74 oThrValGlnLeuHisProSerLeuProLeuProProArgProProArgH 91
|||||
65056 TACAGTCAGCTCCACCCAGGTCTGCCCTCTCCCTCCCGGCCGCCAAGAC 65105
91 isProGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThrSer 107
|||||
65106 ATCCAGAGCCCTCTCTCCCGCGGGGCCAGGGTCCAGAGACCCAGACTTCT 65155

```

```

108 ProGluThrLeuProLeuTyrArgThrGlnSerArgGlyArgGlyGlyPro 124
|||||
65156 CCAGAAACCTCCCTTGTTACAGGACACAGTCTCGGGGAAGGGTGGCC 65205
124 OLeuArgGlyProAlaSerHisLeuGlyArgGluGlnThrGlnIleA 141
|||||
65206 ACTTCAGGTCCCGCTTCCACCTAGGGAGAGAGGAGACCCAGGAGATC 65255
141 rGalaAlaArg..... 144
65256 GAGCGCCAGAGGTGAGAGGCTGGGTGGAAGAGTGGGCTTGGCAA 65305
144 ..... 144
65306 GGTGGGGCTTAGCATGAAGAAAGGGAGCACGGTGGCTTCAGAGGC 65355
144 ..... 144
65356 TTGTCTTTGGTCCAGGAAGTAACCTGGGGAACAGGGTCAGAGTGGTC 65405
145 .....Arg.SerArgLeuArgAs 150
65406 AGAGCTGTGCCCCACATCTGAACACCTTCTCAGGTCCCGCTTCGAGA 65455
150 pProIleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuP 167
|||||
65456 CCCCATCAGCCAGGAATGTTCCGTTATGGGAGAGTCCCTTTGCATTGC 65505
167 rOleuHisArgAsnArgHisProArgSerProProArgSerGluLeu 183
|||||
65506 CACTGCACCGGAACCCAGCAGCACCTCGGAGCCACCACAGATCTGAGCTG 65555
184 SerLeuIleSerSerArgGlyGluGlnAlaIleProSerProThrProAr 200
|||||
65556 TCCTGATCTCTTCTAGAGGGGAAGAGGCTATTCCGTCCTCCCTACTCAAG 65605
200 gAlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProT 217
|||||
65606 AGCAGACCCATTCTCCGAAACGGCAGCCGCCAACTGAGCTCCCTCCCA 65655
217 hrGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerPro 233
|||||
65656 CAGAACTGTGTCTCCACACCCCTTCCGCCCAAGCAGAACTCTAAGCCCT 65705
234 GluThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuAr 250
|||||
65706 GAAACTGCTCAGACAGAGGTGGCCCCCAGAACCCAGGCTGC.CCCCTACG 65754
250 gHisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrH 267
|||||
65755 GCATCACCAGAGAGCCAGGCTGTGSCACAGAGCCGCCCTCACCACGC 65804
267 lSerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArgArg 283
|||||
65805 ACTCCTTAGGAGAAGTGGCTTCTTCGTGATCCCTCAGCCACGAAAG 65854
284 ProSerSerGlnGlyTyrAlaSerProGlnValAlaGlyArgArgProAs 300
|||||
65855 CCAAGTTCACAGGTTGGGCAGTCCGCCAGGTAGCAGGGAGACGCCCTGA 65904
300 pProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProT 317
|||||
65905 TCCCTTTCCTTCGCTCCCTCGGGGCCAGGCCAGCA.GGGCAAGGGCTT 65953
317 rpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspProGln 333
|||||
65954 GGGGAACGGGGGACTCCTCACGGGCCCGCTGGAGCCTGACCTCAG 66003
334 HisProGlyAlaTyrLeuProLeuLeuSerAsnGlyProHisAlaSerSe 350
|||||
66004 CACCCGGCGCTGGCTCCCTGCTGAGCAACGGCCCCCAATGCCAGCTC 66053
350 rLeuTyrSerLeuPheAlaProSerSerProIleProArgCysSerGlyG 367
|||||
66054 CCTCGAGCCCTCTTTGCTCCAGTAGCCCTATTCCCAAGATGTTCTGGG 66103
367 luSerGluGlnLeuArgAlaCysSerGln..Ala..... 377
|||||
66104 AGAGTGAACACTAAGAGCCTGCAGCCCAAGGGTGAGTCTCCTCGGGCCT 66153
377 ..... 377
66154 CCCCTCCCAACCCGACCTCCAGTGTGGCTTCCCTGCCCTGAAGCTGGGT 66203
377 ..... 377
66204 CTTCAGCTTCGCTTGGCACCTGAGGGAGAAGGCCCTTGGCATCTGACCA 66253
377 ..... 377
66254 CCTCAGGCGAGGGGTCTTGGAGCTTTCCTCGCTGACCTGAGCCTCCCACCT 66303
377 ..... 377
66304 GCTCCCCAGGTTACGCCCTGCCCTACCTCATTTGCTCCCCAGCTCG 66353
378 .....ProCysProGluGlnProAspPr 386
66354 ACTCCTTTGTACCCCTCACCGCAGCCCTGCCCCCTGAGCAGCCAGACCC 66403
386 oArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGlnPheMetGlyGlnL 403
|||||
66404 CCGGCCCTTCAGTGGCAGCCTTTAACTCCAGGAATTCATGGGCCACG 66453
403 eutyArgLnrTrpGluProPheThrGluValGlnGlySerGlnArgCysGlu 419
|||||
66454 TGTATCAGTGGGAGCCCTCACTGAA.....GGTGAG 66485
420 LeuAsnCysArgProArgGlyPheArgPheTyrValArgHisThrGluLy 436
|||||
66486 GTTCTTGTCTACCCCTGGC..... 66506
436 sValGlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysVala 453
|||||
66507 .....AGTGGTGGCTTGGAAATTGGGGATGA 66531
453 laGlyArg.....CysLeuSerProGlyCysAspGlyIle..... 464
66532 AGGGGAGAGGAGATACCTGCTTACTCCAGCCCTGAATGACTTCCAGCC 66581
465 .....LeuGlySerGlyAr 469
66582 CTCTGCTTCCCTGGCCATGCCCTTCTTCTTCTCCCTGGGCTGGGTCA 66631
469 gArgProAspGlyCysGlyValCysGlyGlyAspAspSerThrCysArgL 486
66632 G.....GAGACAGCACAGG 66645
486 euValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLys 502
66646 TGGTGAAGTGTGTCAGAGGGTGGGGTGGG..... 66677
503 ileLeuTyrIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuAr 519
66678 .....AG 66679
519 gProSerSerAspTyrLeuAlaLeuArgGly.ProGlyGlyArgSerIle 535
66680 GAGAGGTGTGGAGACACGCTTTTGTCCGACTCCCTCGGAAGGCCATC 66729
536 ileAsnGlyAsnTrpAlaVal.....AspPr 544
66730 ACTGGGGTCAAGTGGCTGTGACACAAGAGGCCATGTTGGGAGAGATCC 66779
544 oProGlySerTyrAlaGlyGlyThrValPheArgTyrAsnArgProp 561
|||||

```

103 TGGGACACTCTGCATGCTCAGCGTCCTGCGGGAAGGTGCTGGCGCCCC 152



810 ysValGlyAsnGlyAspGluValSerGluGlnGluCysAlaSerGly 826  
 1220 GTGTGAGCAACATTGGGATGGTGTGACGATGAGGAATGC...AACATG 1266  
 827 ProProGlnProProSerArgGluAlaCysAspMetGlyProCysThrTh 843  
 1267 AACCTCCGGCCGAATGACATTCAGACTCGACATGGGACCTGTGCCA 1316  
 843 rAlaTrpPheHisSerAspTrpSerSerLysValSerProGlu 857  
 1317 GAGCTGTTCTCACCAGTGGAGCGAAAGTGTCTCAGCGGAG 1359

seq\_name: gb\_pat:E55265

seq\_documentation\_block: 5061 bp DNA linear PAT 31-JAN-2002  
 LOCUS E55265  
 DEFINITION Novel metalloprotease and gene of the same.  
 ACCESSION E55265  
 VERSION E55265.1 GI:18629778  
 KEYWORDS JP 2001008687-A/1.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.

REFERENCE  
 1 (bases 1 to 5061)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AUTHORS Yamaji,N., Nishimura,K. and Sasamata,M.  
 TITLE Novel metalloprotease and gene of the same  
 JOURNAL Patent: JP 2001008687-A 1 16-JAN-2001;  
 YAMANOUCHI PHARMACEUT CO LTD

COMMENT  
 OS Homo sapiens (human)  
 PN JP 2001008687-A/1  
 PD 16-JAN-2001  
 PF 25-JUN-1999 JP 1999180973  
 PR  
 PI NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA  
 PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC  
 C12N9/64, C12Q1/37,  
 PC C12N15/00, C12N5/00

CC  
 FH Key Location/Qualifiers  
 FT source 1..5061  
 FT /organism='Homo sapiens (human)'.  
 FT Location/Qualifiers

FEATURES  
 source 1..5061  
 BASE COUNT 878 a 1746 c 1587 g 850 t  
 ORIGIN

alignment\_scores:  
 Quality: 807.50 Length: 959  
 Ratio: 1.815 Gaps: 41  
 Percent Similarity: 46.403 Percent Identity: 28.050

alignment\_block:

US-10-041-770-2 x E55265 ..

Align seg 1/1 to: E55265 from: 1 to: 5061

15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24  
 :|||||: |||  
 596 GTGCTCCAAGCACTGTGGAGTGCAGTGTACCAGAGCTGGAGTCTCGA 645  
 24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35  
 :|||||: |||  
 646 CGGAGCGGTGGGAGCAGCGGACAGTGTCCGCGCGCGGCTGAGGGC 695  
 35 hrProThrGluGluGlnGlyProGluGlyValTrpGlyProTrp... 50  
 :|||||: :|||  
 696 TCTACACCGGGTCTGCTGACCAAGAGAGTGGTGGAGACCTGTGTAG 745  
 50 .....

746 TAGTGATGCCAAATAGTGGAGTACCAGGACAGCGCAGGTTGAGAGC 795  
 51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61  
 :|||||: |||  
 796 TATGTGTGACCATCATGAACATGGTGGCTGTTTCATGACCCAG 845  
 61 lyValGlyVal.....Gln 65  
 :|||||: :|||  
 846 CATTGGGAACCCCATCCACATCACCATTGTGGCTGCTGCTGGAAG 895  
 66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80  
 :|||||: :|||  
 896 ATGAGGAGGAGGACCTAAAGATCAGCCACCATCAGACACACCCCTGAAG 945  
 81 .....SerLeuProLeuP 85  
 :|||||: :|||  
 946 AGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAAGGGGTGCCATCC 995  
 85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
 :||  
 996 CCTGCACCATG..... 1006  
 102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118  
 :|||||: |||  
 1007 ACATGTCATCTGCTCACCAGAA.....AGGACCTGTGT 1041  
 118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135  
 :|||||: :|||  
 1042 CGAGCCATGAACCGGCTGTGAGACCTGGGACTGCCATGTGGCGGG 1091  
 135 luGluThrGlnGluLeuAlaAlaArgArgSerArgLeuArgAspPro 151  
 :|||||: :|||  
 1092 CATGTGCCAGCGCCAGCCAGTCAGTCATCAGCAGGAGGACA..... 1132  
 152 IleLysProGlyMetPheGlyTyGlyArgValProPheAlaLeuProLe 168  
 1132 ..... 1132  
 168 uHisArgAsnArgHisProArgSerProProArgSerGluLeuSerL 185  
 :|||||: |||  
 1133 ...CGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1178  
 185 euLeuSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201  
 :|||||: |||  
 1179 TTTTGGCATGACATGACGGAAGCGGCAATGACTGTGAGCCCTGTGGGA 1228  
 202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro.. 215  
 :|||||: :|||  
 1229 AACGACCTTCATCATGTCTCCACAGCTCTGTACGACCGGCTCCGCTC 1278  
 216 ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232  
 :|||||: :|||  
 1279 ACCTGTCCCGCTGCGCGGCGGATATATCACCAGTTCTTGTGACCGTGG 1328  
 232 exProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248  
 :|||||: :|||  
 1329 GTGGGGCTGTGCTGGAGCAGCCCTCTGCTGCCAAGGACATTATCGACTCC 1378  
 249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258  
 :|||||: |||  
 1379 COTCGTCCGACCTGCTGCTCTCTATGATGTAAGCCACCATGTCGCTC 1428  
 258 rGlyThrGluProProSerPro.....T 266  
 :|||||: |||  
 1429 CAGTACGGGGCTACTCTGCCTCTCTGCGGAGGACATGGATATGTCTGCCA 1478  
 266 hrHisSerLeuGlyGlyGlyPhePheArgAlaSerProGlnProArg 282  
 :|||||: :|||  
 1479 CACACTGTGCTGTGTGGGA.....CCACCTGTC 1510  
 283 ArgProSer....SerGlnGlyTrpAlaSerProGlnValAlaGlyArg... 297  
 :|||||: |||

1511 ACTCAAGCTGGATGACGGCTGGACGGCACCCGGTGTG...GGGAGAAAT 1557  
298 .....ArgProAspProPheProSerValProArgG 308  
1558 AAGTGTGTCTAGTGGGAGATGCGTACCGTGGGCTTCCGGCCCGAGG. 1606  
308 lyArgGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321  
1607 .....CCGTGGATGTGGCTGTGGTGGTGA 1633  
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338  
1634 GGCCT...GGTCCA..... 1645  
338 pLeuProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355  
1646 .....TCTGCTCAGCGAGCTGTGGCATGGCGGTACA 1676  
355 heAlaProSer...Ser.....ProLeuProArg..... 363  
1677 GAGCGCGAGCGGAGTGCACGAGCTACGCCCAATAATACAAAGGCAGAT 1726  
364 ..CysSerGlyGluSerGluGlnLeuAlaCysSerGlnAlaProCys 379  
1727 ACTGTGTGGTGAGCGCAAGCGCTTCGGCTCTGCAACCTGCAGGCGCTGC 1776  
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSe 396  
1777 CTTGCTGGCCACCCCTCTTCCGCCACGTCCTCAGTGCAGCCACCTTTCACGC 1826  
396 rGlnGluPheMetGlyGlnLeuTrpGlnTrpGluProPheThrGluValG 413  
1827 TATGCTCTACAAGGCCAGCTGCACATCGGTGCCCGTG.....GTCA 1870  
413 lnGlySerGlnArgCysGluLeuAsnCysArgProArgProGlyPheArgPhe 429  
1871 ATGACGTGAACCCCTGCGAGCTGCACTGCGCGCGCGCAATGAGTACTTT 1920  
430 TyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro... 445  
1921 GCCAGAGAGTCGGGACCGCGTGTGTCATGGCACCCCTGCTACCAAGGT 1970  
446 ....GlyAlaProAspLeuCysValAlaGlyArgCysLeuSerProGlyC 461  
1971 CCGAGCCAGCGGGACCTCTGTCATCAACGCACTCTGTAAGACGTGGCT 2020  
461 ysAspGlyLeuLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477  
2021 GTGACTTCGAGATTGACTCCGGTGTCTATGAGGACCGCTGTGGTGTGTC 2070  
478 GlyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspAr 494  
2071 CACGGCAACGGCTCCACTGCCACACCGTGAGCGGACCTTCGAGGAGGC 2120  
494 gGlyGlyProLeuGlyThrGlnLysIleLeuTrpIleProAlaGlyAlaL 511  
2121 CGAGGCG...CTGGGGTATGTGGATGTGGGCTGATCCACGCGCGCAC 2167  
511 euArgLeuGlnIleAlaGlnLeuArgProSerAsnTrpLeuAlaLeu 527  
2168 GCGAGATCCGATCAAGAGGTGCGGAGGCTGCCACTTCCTGGCACTG 2217  
528 ArgGlyProGly...GlyArgSerIleIleAsnGlyAsnTrpAlaValas 543  
2218 CGGAGTGAGGACCGGAGAGTACTTCTCAATGGTGGTGGACCATCCA 2267  
543 pProProGlySerThrArgAlaGlyGlyThrValPheArgTrpAsnArgP 560  
2268 GTGGAACGGGAGTACCAAGGTGGGAGGACCACTTCATCATCCACGCG. 2316  
560 roProArgGluGlyLysGlyGluSerLeuSerAlaGluGlyProThr 576  
2317 .....AGGGCAACTGGGAGAACCTCACGTCCCGGGTCCCAAC 2355

577 ThrGlnProValAspValTyrMetIlePheGlnGluAsnProGlyVa 593  
2356 AAGGAGCCTGTGTGATCCAGCTGCTTCCAGGAGCAACCTGGGGT 2405  
593 lPheTyrGlnTrpValIleSerSerProProIleLeuGluAsnProT 610  
2406 GCATACAGTAC..... 2418  
610 hrProGluProProValProGlnLeuGlnProGluIleLeuArgValGlu 626  
2418 ..... 2418  
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2419 .....ACCATCCA 2426  
643 nArglnValArgIleProGlnMetProAlaProProHisProArgThrP 660  
2427 CAGGGAG.....GCAGGTGGCCACGACGAGGTCC 2455  
660 roLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCys 676  
2456 CG.....CCGCCGTGTCTCTGGCATTATGGCCCTGGACCAAGTGC 2499  
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693  
2500 ACAGTCACCTGCGGAGAGGTGTGAGGAGGCAATGTGTACTGTGGA 2549  
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710  
2550 GCGGACGAGGAGG...CCGTGGAGGAGGACACTGTGACCCCTGGGCC 2596  
710 rgProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726  
2597 GGCCTGATGACCAACAGAGGAAGTGCAGGAGCAGCCCTGCCCTGCCAGG 2646  
727 TrpGluAlaGlyLclTrpThrSerCysSerArgSerCysGlyPro...G 742  
2647 TGTGGGCAAGTGTAGTGGCAGCTGTCTCCAGCTCTCGGGGCTGGGG 2696  
742 TyrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGlyG 758  
2697 CTTCTCCCGCGGGCGGTCTGTGATCCGACGCTGGGCTGGATGAGC 2746  
758 lySerSerValProProGluArgCysGlyHisLeuProArgProAsnIle 774  
2747 AGAGCGCCTGGAGCCACCCGCTGTGAACACTTCCCGCGCCCTACT 2796  
775 ThrGlnSerCysGlnLeuArgLeu...Cys....GlyHisTrpGluValG 789  
2797 GAAACCCCTGCAACCGCATGTACCTGTCCGCGCCACCTGGGCTGGG 2846  
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerA 806  
2847 GAAC...TGGTCTCAGTGTCTAGTGCATGTGGGAGGACACTCAGCGCC 2893  
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822  
2894 GAAATGTCTCTGCACCAATGACACCGGTGCCCTGTGACGAGGCCAG 2943  
823 CysAlaSerGlyProProGlnProProSerArgGluAlaCysAspMetG 839  
2944 .....CAGCCAGCCAGCAAGTACCTGCTCTCTGCC 2975  
839 yProCysThr...ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854  
2976 ACTCTGTGGTGGCCCTGGGCACAC.....TGGGCCCTGAAGGCTCAG 3019  
855 SerProGluProProAlaIleSer 862  
3020 GCAGCGGCTCTCTCAGCCAGCAGC 3043





258 rGlyThrGluProSerPro.....T 266  
1464 CAGTACGGGCGCTACTCTCTCTCGAGGACATGATAATGTCTGCCA 1513  
266 hrHisSerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArg 282  
1514 CACACTCTGTGTCTGTGGGA.....CCACTGTC 1545  
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg.. 297  
1546 ACTCAAGCTGGATGACGCGGTGGAGCGACCCGGTGTG...GGGAGAA 1592  
298 .....ArgProSerProPheProSerValProArg 308  
1593 AAGTGTGTCTCAGTGGGAGTCCGACCGTGGGTTCGGGCCGAGG.. 1641  
308 lyArgGlyGlnGlyGlnGlyProTrp.....GlyThrGlyGly 321  
1542 .....CCGTGGATGTGTGTGTGTGTGGA 1668  
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338  
1669 CGGCT...GGTCCA..... 1680  
338 pLeuProLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355  
1681 .....TCTGCTACGGAGCTGTGGCATGGCGTACA 1711  
355 heAlaProSer..Ser.....ProIleProArg..... 363  
1712 GAGCGCGAGGCGAGTCACGCGAGCTAGCGCCCAATACAAAGCGAGAT 1761  
364 ..CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCys 379  
1762 ACTGTGTGGTGACGGAAGCGCTTCGGCTCTCGCACTCGACCTGCGGCTGC 1811  
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSe 396  
1812 CTGTGTGCGCCCTCTCTCCGCCAGCTCCAGTGCAGCCACTTTCAGC 1861  
396 rGlnGluPheMetGlyGlnLeuTrpGlnTrpGluProPheThrGluValG 413  
1862 TATGCTCTACAGGCGCAGCTGCAATGGGTCCCGTG.....GTCA 1905  
413 lnglySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhe 429  
1906 ATGAGCTCAACCCCTCGAGCTGCACTGCGCGCCGCGAATAGTACTTT 1955  
430 TyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro.. 445  
1956 GCCGAGAAGCTCGGAGCGCGTGTGTGATGGCACCCCTCTACAGGT 2005  
446 ....GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyC 461  
2006 CCGAGCCAGCGGACCTCTCATCAAGGCATCTGAAGAAGCTGGGCT 2055  
461 ysAspGlyLeuLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477  
2056 CTGACTTCAGATGACTCCGCTCTATGGAGGACCGCTGTGTGTGTGTC 2105  
478 GlyGlyAspSerThrCysArgLeuValSerGlyAsnLeuThrAspAr 494  
2106 CAGGCAAGGCTCCACTGCCACACCGTGAAGCGGACCTTCGAGGAGGC 2155  
494 gGlyGlyProLeuGlyThrGlnLysIleLeuTrpIleProAlaGlyAlaL 511  
2156 CGAGGCG...CTGGGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2202  
511 euArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTrpLeuAlaLeu 527  
2203 GCGAGATCGCATCCAGAGGTTCGCGAGGTGCGCACTTCTGCGACTG 2252  
528 ArgGlyProGly...GlyArgSerIleAsnGlyAsnTrpAlaValAs 543

2253 CGGAGTGAAGACCGGAGAGTACTCTCTCAATGGTGGCTGGACCATCCA 2302  
543 pProProGlySerTyArgAlaGlyGlyThrValPheArgTyArgAsnArgP 560  
2303 GTGGAAACGGGACTTACCAGTGGCAGGACCACTTCCATACACGACGC. 2351  
560 roProArgGluGluGlyLysGlyGluSerLeuSerAlaGluGlyProThr 576  
2352 .....AGGGCACTGGGAGAACCTCACGTCCCGGGTCCCACC 2390  
577 ThrGlnProValAspValTyMetIlePheGlnGluGluAsnProGlyVa 593  
2391 AAGGACCTGTCTGGATCCAGCTGTGTTCAGAGAGAGCAACCTGGGT 2440  
593 lPheTyGlnTyTrpValIleSerSerProProIleLeuGluAsnProT 610  
2441 GCACACGAGTAC..... 2453  
610 hrProGluProProValProGlnLeuGlnProGluIleLeuArgValGlu 626  
2453 ..... 2453  
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGl 643  
2454 .....ACCATCCA 2461  
643 nArgGlnValArgIleProGlnMetProAlaProHisProArgThrP 660  
2462 CAGGAG.....GCAGTGGCCACGACGAGGTCC 2490  
660 roLeuGlySerProAlaAlaTyTrpLysArgValGlyHisSerAlaCys 576  
2491 CG.....CCGCCGTGTCTCTGGCATATGGGCCCTGGACCAAGTGC 2534  
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693  
2535 ACAGTCACTCGGCGAGAGTGTGCAGAGGACAGAAATGTACTGTCTGGA 2584  
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710  
2585 CGGCGAGGAGG...CCGCTGACGAGGAGCACTGTGACCCCTGGGCC 2631  
710 rProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726  
2632 GGCTGATGACCAACAGAGAGTGCAGGACGAGCCCTGCCCTGCCAGG 2681  
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...Gl 742  
2682 TGCTGGCAGTGAAGTGGCAGCTGTCTCAGCTCTCGGCGCTGGGGG 2731  
742 yThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGlyG 758  
2732 CACTCCCGCGCGCGTGTCTGTCTGATCCGACGCTGGGCTGGATGAGC 2781  
758 lySerSerValProGluArgCysGlyHisLeuProArgProAsnIle 774  
2782 AGAGCCCTGGAGCCACCGCTGTGAACACCTTCCCGGCCCTACT 2831  
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValG 789  
2832 GAAACCCCTTGAACCGCCATGATACCTTCCCGGCCACCTGGGCTGTGG 2881  
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerA 806  
2882 GAAC...TGCTCAGTGTCTGATGATGTGGGAGGCACTCAGCGCC 2928  
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822  
2929 GAAATGTCTCTGACCAATGACACCGGTGTCCCTGTGACGAGGCCAG 2978  
823 CysAlaSerGlyProGlnProProSerArgGluAlaCysAspMetGl 839

2979 .....CAGCCAGCCAGCGAAGTCACCTGCTCTCTGCC 3010  
 839 yProCysThr.ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854  
 3011 ACTCTGCTGGTGGCCCTGGGCACAC.....TGGCCCTGAGGCTCAG 3054  
 855 SerProGluProProAlaIleSer 862  
 3055 GCAGCGGCTCCTCCAGCCACGAGC 3078

seq\_name: gb\_pat:AX319851

seq\_documentation\_block:  
 LOCUS AX319851 3312 bp DNA linear PAT 14-DEC-2001  
 DEFINITION Sequence 15 from Patent WO0183782.  
 ACCESSION AX319851  
 VERSION AX319851.1 GI:17901441  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (sites)  
 AUTHORS Plowman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and  
 Payne,V.  
 TITLE Novel proteases  
 JOURNAL Patent: WO 0183782-A 15 08-NOV-2001;  
 Sugen, Inc. (US)  
 FEATURES  
 Location/Qualifiers  
 1..3312  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 620 a 1106 c 1035 g 551 t  
 ORIGIN

alignment\_scores:  
 Quality: 797.50 Length: 954  
 Ratio: 1.693 Gaps: 45  
 Percent Similarity: 49.371 Percent Identity: 28.512

alignment\_block:

US-10-041-770-2 x AX319851 ..  
 Align seg 1/1 to: AX319851 from: 1 to: 3312  
 4 TrpThrGlyArgProTrpLeuLeuLeuLeuSerLeuProGlu 20  
 604 TGAAGAGCGCGCCATGG...TGCTGGGACCTTGAAGCCACCGCTGC 650  
 20 nLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuGlnThrPro 37  
 651 CAGGCCCTGGGAATGAA.....A 670  
 37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46  
 671 CAGAGCGTGGCCAGCCAGGCTGAAGCGATCGTCAGCGAGAGCGCTAC 720  
 46 lTrpGlyProTrp..... 50  
 721 GTGGAGACCTGTGGTGGTGACAGATGATGGTGGCCTATCACGGGCG 770  
 51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66  
 771 CCGGGATGTGGAGCAGTATGCTCTGCCGTCATGAACATTTGGCCAAAC 820  
 67 ArgSerArgThr..... 70  
 821 TTTTCCAGGACTCGAGTCTGGGAGCACCCTTAACATCCTCGTAATCGC 870  
 71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85  
 871 CTCATCCTGCTCAGGAGGAGGCCACCTCTGGAGATCA..... 910

85 roProA-gProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
 911 ..CCACCATCGCGGAAGTCCCTGGACAGCTTCTGTAAGTGGCAGAAAT 958  
 102 ProArgProGlnThrSerProGluThrLeuLeuProLeuTyArgThr 116  
 959 CCATCGTGACCAACAGCGCCATGGCAATGCCATTCAGAGACGCGTGTG 1008  
 117 .....GlnSerArgGlyArgGlyGlyP 124  
 1009 GCTAACCATGACACAGCAGTCTCATCACAGCGTATGACATCTGCATCTA 1058  
 124 roLeuArgGlyProAlaSerHisLeuGlyArgGluThrGlnGluLe 140  
 1059 CAAGAACAACCCCTGGCGCACACTAGCCCTGGCCCGGTGGCGGGAATGT 1108  
 141 ArgAlaAlaArgArg.....SerArgLeuArgAspProIleLysProG 155  
 1109 GTGAGCGCGAGAGAGCTGCAGCGTCAATGAGGACATTCGGCTGGCCACA 1158  
 155 yMetPheGlyTyArgValProPheAlaLeuProLeu.HisArgAsn 171  
 1159 G.....CGTTCACCATTCGCCACGAGAT 1181  
 172 ArgArgHisProArgSerProProArgSerGluLeuSerLeuLeuSer 188  
 1182 CGGCACACATTCGCATGAACCA.....TGACG 1210  
 188 rArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPhe 205  
 1211 GCGTGGGAACAGCTGTGG.....GGCCCGTGTGTTCAGGACCCAGCC 1251  
 205 eAlaAsnGlySerProGln...ThrGluLeuProProThrGluLeuSer 220  
 1252 AGCTCATGGTGGCCCATACCATGACATGAAGACCAACCATTCGGTGTGTC 1301  
 221 ValHisThrPro.....SerProGlnAlaGluProLeuSerProGluTh 235  
 1302 ATCCTGACGCGCTGACTACATCACAGCTTTCTAGACTCGGCGCTGGGC 1351  
 235 rAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArgHis 252  
 1352 TCTGCTGAACAACCGCGCCCGCAGACAGGACTTGT..... 1388  
 252 isProArgAlaGlnAlaSerGlyThrGluPro..... 262  
 1389 .....GTACCCGACAGTGGCACCAGCCAGCCTACGATGCAGATGA 1430  
 263 ProSerProThrHisSerLeuGlyGlu.....G 272  
 1431 GCAATGCCGCTTTCAGCATGGATCAATCGCTCAGTGAATACGGGG 1480  
 272 yGlyPhePheArgAla.....SerProGlnPro..... 281  
 1481 AGTCTGCAGCGAGTGTGGTGTCTGACGACGAGCAACCCGTGCATCAC 1530  
 282 ..ArgArgProSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArg 297  
 1531 AACAGCATCCCGCGCGCGAGGCGACGCTGTGCCAGACGACACCATCGA 1580  
 298 ArgPro.....AspProPheProSerValProAr 307  
 1581 CAAGGGGTGCTGTACAAACGGGTCTGTCTCCCTTGTGGTGGCGCCAG 1630  
 307 gGlyArgGlyGlnGlyGlnGlyProTrpGlyThrGlyThrProH 324  
 1631 AGGGGTGGACGG.....AGCCTGGGGCGCGTGGACT 1662  
 324 isGlyProArgLeuGluProAspProGlnHisProGlyAlaTrpLeuPro 340  
 1663 CCATGGGGCGACTGCAGCGGAC..... 1685

341 LeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSer.....Le 354  
1686 .....CTGTGGCGGCGGTCTCT 1705  
354 uPeAlaProSer.SerProIleProArg..... 363  
1706 CTCTACTCGGTCTACTCGACAGCCCGACGCCAACCATCCGGGGGCAAGTAC 1755  
364 CysSerGlyLeuSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380  
1756 TGTCTGGGTGAGAGAGCGGCGCACCCCTCTGCAACACCGATGATGTCTC 1805  
380 OProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397  
1806 CCCTGGCTCCAGGACTTCAGAGAAGTGCAGTCTTCTGAATTGACAGCA 1855  
397 LcLuphMetGlyGlnLeuTyrlGlnTrpGluProPheThrGluValGln 413  
1856 TCCTTTCCGTGGGAATCTCAAGTGAACACGTAC...CGGGAGGG 1902  
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTy 430  
1903 GCGGTGAAGCCCTGCTCGCTACGTGCGTACGGAAGGCTTCAACTTCTA 1952  
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447  
1953 CACGGAGAGGGCGGCAAGCCGTGTGACGGGACACCCCTGCCGTCCAGACA 2002  
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463  
2003 CGGTGGACATTTGGTTCAGTGGGAGGAGTCAAGCACCGTGGCTCGACCGA 2052  
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAs 480  
2053 GTCTGGGCTCCAGCTCGCGGAGGACAAAGTCCGAGTGTGTGGCGGTGA 2102  
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497  
2103 CGGAGTGCCTGGAGACCATCAGGCGCTTCAGCCCGACCTCAGCTG 2152  
497 roLeuGlyTyrlGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeu 513  
2153 GGGCGGGTACGAGGATGTGTGGATTCCCAAGGCTCCGTCCACATC 2202  
514 GlnIleAlaGlnLeuArgProSerSerAsnTyrlLeuAlaLeuArgGlyPr 530  
2203 TTCATCAGGATCGAACCTCTCTCAGTCACTTGGCCCTGAAGGAGA 2252  
530 oGlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspProProGlys 547  
2253 CCAGGACTCCCTGCTGTGGAGGGCTGCCGGGACCCCGACCCCGACCC 2302  
547 eTyrlArgAlaGlyThrValPheArgTyrlAsnArgProProArgGlu 563  
2303 GTCTGCCTCTAGCTGGGACCACTTTCACCTGGACAGGGCGGACGAC 2352  
564 GluGlyLysGlyGluSerAlaGluGlyProThrThrGlnProva 580  
2353 GTC.....CAGAGCTCGAAGCCCTGGGACCGATTAAATGATCTCT 2393  
580 lAspValTyrlMetIlePheGlnGluGlnAsnProGlyValPheTyrlGlnT 597  
2394 CATGCTCATGTGTGTGGCCCGGACCGAGCTGCTGCTGCTGCTGCTGCT 2443  
597 yrValIleSerSer.ProProIleLeuGluGlnAsnProThrProGluPr 613  
2444 TC.....AATGCCCGCATCGCGTGTGCTGCTGCTGCTGCTGCTGCT 2484  
613 o.....ProValProGlnLeuGlnProGlu..... 621  
2485 TGGCACTATGGCCCTTGACCAAGTGTCTGGCCCGGACGAGTGTGACGCGGTAG 2534  
622 .....Ile.LeuArgValG1 626

2535 CCAGGTGCAGCGGTGGAGTCCGCAACACAGCTGGACAGCTCCCGGTGCG 2584  
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2585 CCCCCCACTACTGTCAGTCCACACAGAGCT.....GCC 2619  
643 lNArgGlnValArgIleProGln.MetProAlaProProHisProArgTh 659  
2620 AAAAGGCAGCGCTGCAACACAGGAGCTTGCCTCCAGAC..... 2661  
659 rProLeuGlySerProAlaAlaTyrlTrpLysArgValGlyHisSerAlaC 676  
2662 .....TGGTGTAGGAACCTGGTCTGCTCT 2686  
676 ysSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIle 692  
2687 GCAGCGCGAGCTGCGATCGAGCGTGCAGCGCTCGTGTGTGTGCGAC 2736  
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706  
2737 CGCCCGCTCTCTGCGCGGAGAGAGCGCTGGACGACGACGCGATGCC 2786  
706 aAlaGlyAlaArgProAlaSerProGluProCysHisGlyThrProC 723  
2787 GCAG...CGCGCCACCTGTACTG...GAGCCTGCCACGCGCCACTT 2830  
723 ysProProTyrlTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739  
2831 GCGCTCCGAGTGGCGCGCTCGACTGTGTGTGTGTGTGTGTGTGTGTGT 2880  
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyG1 756  
2881 GGGCGGCGCTCCGCGACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2924  
756 yGlyGlySerSerValProProGluArgCysGlyHisLeuProArgProA 773  
2925 CCACGCGCGACGCTGCCCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGT 2974  
773 snIleThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787  
2975 CGGCGCACCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3024  
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804  
3025 GCTGGCGAG...TGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3071  
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821  
3072 GAGCGCTCGGTGCGCTGCACGACGACGACGAGGCGCAG.....GCGTGC 3115  
821 lNlGluCysAlaSerGlyProProGlnProProSerArgGluAlaCysAsp 837  
3116 ACGAGTGCAGGAGGCC...CTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3162

seq\_name: gb\_pat:E55273

seq\_documentation\_block: 3312 bp DNA linear PAT 31-JAN-2002  
LOCUS E55273  
DEFINITION Novel metalloprotease and gene of the same.

ACCESSION E55273  
VERSION E55273.1 GI:18629786  
KEYWORDS JP 2001008687-A/9.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Yamaji,N., Nishimura,K. and Sasamata,M.  
TITLE Novel metalloprotease and gene of the same  
JOURNAL Patent: JP 2001008687-A 9 16-JAN-2001;  
YAMANOUCHI PHARMACEUT CO LTD  
COMMENT OS Homo sapiens (human)



```
397 lnGluPheMetGlyGlnLeuTyrGlnTrpGluProPheThrGluValGln 413
1856 TCCTTTTCCTGGGAAATCTACAAAGTGAACACGTAC...CGGGAGGG 1902
414 GlySerGlnArgCysGlnLeuAnCysArgProArgGlyPheArgPheTy 430
1903 GGGTGAAGCCCTGCTCGCTACGCTAGCGGAAGGCTTCAACTCTA 1952
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447
1953 CACGGAGAGGGCGGACCGCTGTGTGACGGGACACCTCCGTCACAGACA 2002
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463
2003 CGGTGGACATTTCGTCAGTGGCGAATGAAGCACCGTGGCTCGACCGA 2052
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyLys 480
2053 GTCTGGGCTCCGACCTCGGGAGGAGCAAGTCCGAGTGTGTGGCGTGA 2102
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497
2103 CGGCAGTGCCTGGAGACCATCGAGGGCTCTTCAGCCCGACCTCACCTG 2152
497 roLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeu 513
2153 GGGCGGGTACGAGGATGTGCTGTGATTCCCAAGGCTCCGTCACATC 2202
514 GlnIleAlaGlnLeuArgProSerAsnTyrLeuAlaLeuArgGlyPr 530
2203 TTCACAGAGCTGAACCTCTCTCAGTCACCTTGGCCCTGAAGGAGA 2252
530 oglyGlyArgSerIleLeuAsnGlyAsnTrpAlaValAspProGlyLys 547
2253 CCAGGAGTCCCTGTGTGTGGAGGGCTGCGGGAGACCCCGCCAGCCACC 2302
547 erTyrArgAlaGlyThrValPheArgTyrAsnArgProProArgGlu 563
2303 GTCTGCCTCTAGCTGGGACCACTTTCACCTGCACAGGGGCCAGACCAG 2352
564 GluGlyLysGlyGlyLeuSerLeuAlaGluGlyProThrTrpGlnProVa 580
2353 GTC.....CAGAGCTCGAAGCCCTGGGACCGATTAAATGATCTCT 2393
580 lAspValTyrMetIlePheGlnGluGlnAsnProGlyValPheTyrGlnT 597
2394 CATGCTCATGTGTGTCGCGGACCGAGCTGCTGCTGCTGCTGCTGCTGCT 2443
597 yrValIleSerSer.ProProIleLeuGluAsnProThrProGluPr 613
2444 TC.....AATGCCCATCGCCGCTGACTCGTGCCCC...CCTACTCC 2484
613 o.....ProValProGlnLeuGlnProGlu..... 621
2485 TGGCACTATGCGCCTGACCAAGTGTGCGCCCAAGTGTGCGCGGTAG 2534
622 .....Ile.LeuArgValG 626
2535 CCAGTGGAGCGGTGGAGTCCCGACACCGAGCTGGACAGCTCCGCGTGC 2584
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643
2585 CCCCCCACTACTGAGTGCACACAGCAAGCT.....GCC 2619
643 lnArgGlnValArgIleProGln.MetProAlaProProHisProArgTh 659
2620 AAAGGACGCGCGCTGCAACAGGAGGCTTGCCTCCAGAC..... 2661
659 rProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaC 676
2662 .....TGGGTGTAGGGAAGTGGTGGCTCT 2686
676 ysSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIle 692
```

```
2687 GCAGCCGCGAGCTGGATCGAGCGTGGCGCGCTGGTGGTGGCCAG 2736
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706
2737 GCGCGCTCTCTGCGCGGAGGAGGAGCGCTGGAGACGACGCGATGCC 2786
706 aAlaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProC 723
2787 GCAG...CCGCGCCACCTGTACTG...GAGCGCTGCCACGGCCCACTT 2830
723 ysProProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739
2831 GCGCTCCGAGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2880
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyG 756
2881 GGGCGGGGCTCCGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2924
756 yGlyGlySerValProProGluArgCysGlyHisLeuProArgProA 773
2925 CCACCGCGCAGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 2974
773 snIleThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787
2975 CGCGCCACCATGCGCTGCACTTTCGCGCTGCGCGCGCGCTGCGGTG 3024
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804
3025 GGTGGCGAG...TGGGGTGAAGTCTGTCACAGTGGCGCTGCGGCGAG 3071
804 gSerArgGlnValArgCysValGlyAsnGlyAspGluValSerGluG 821
3072 GAGCGCTCGGTGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCG 3115
821 lnGluCysAlaSerGlyProProGlnProProSerArgGluAlaCysAsp 837
3116 ACGAGTGCAGGAGGCC...CTGCGCGCGCGCGCGCGCGCGCGCGCG 3162
```

seq\_name: gb\_pr:AF163762

seq\_documentation\_block:  
LOCUS AF163762 3400 bp mRNA linear PRI 01-DEC-2000  
DEFINITION Homo sapiens zinc metalloendopeptidase (ADAMTS10) mRNA, partial cds.  
ACCESSION AF163762  
VERSION AF163762.1 GI:11493588  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
ADAM-TS10: A novel member of the ADAM-15 family containing multiple thrombospondin type I repeats  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
Direct Submission  
AUTHORS  
JOURNAL  
TITLE  
Submitted (29-JUN-1999) Biomedical Engineering, Lerner Research Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA  
FEATURES  
source  
1..3400  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..3400  
/gene="ADAMTS10"  
/note="a disintegrin-like and metalloprotease domain with thrombospondin type I repeats-10"  
<1..3234  
/gene="ADAMTS10"

```

/codon_start=1
/product="zinc metalloendopeptidase"
/protein_id="AAG35563.1"
/db_xref="GI:11493589"
/translation="RSODELSLESVEIAFPRVDHNGALLAFSPPPRRORRGTA
TAESRLYKVASSTHPLNLTRSSRLLAGHSVVEYTRREGLAWQRAARPHCLYAGHL
QGQASSHVAISCGGLHGLIVADEEYLIEPLHGPKGSRPEESGPHVYKRSILR
HPHDTACGVDERKPKRPWMLNTLKPAPPLGNTERGQGLKARSREIRIVETL
VADKMMVAYHGRDVEQYVLAIMNIVAKLFQDSLSGSTVNIILYLLITLDPQLE
ITHHAGKSLDSFCWKQKSYVHSGHNAIPENGVAHNDTAVLTIRYDICTYKKNPCGT
LAPVGMGERSCSVNEDIGLPQATTAHIGHTFGMHDGVGNSCGARQDPFAK
LMAAHITMKNPVMSSNEDYITSPDLSGLCLNNRPPQDFVPTVAPQAYDAD
EQCFQHGKASROCKYGVCSLWCLSKSNRCITNSTPAEGLTLCOTHTDKGWCKYR
VCFPGRPEVGDWAGPFWDCSRTCCGGYSSSSRHCDSRPRTIGGKYCLGERRR
HRCNTDCCPGSDREVQCSEFDSIPFGKFKWKYIRGGVVKAGSLSLSLAEGNF
YTERAAAVDGTGCPDPTVDICVSGECHKVDCDLREDKRCVCGDGGACETI
EGVFPSPAGAGYEDVYVPIKGSYHIFIQDLNLSHLALKQDSLLLEGLPFGPOP
HRLPLAGTTFOLRGQPDQVOSLEALPINASLIYVLAITELPALRYRFNAPLAROSL
PPYSWHVAPWTKCSAQACAGSOVOAVECNQOLDSAVAPHYCSAHSKLPKRQACNTE
PCPDVWGNLSLCSRCODAGVRSVVCORRVSAAREKALDDSDACPPRPPVLEACH
GPTCPPEWALDNECTPSCGPGLRHVRVICKSADHRATLPPAHCSAAKPPATMRCN
LRPCPPAHVAGENGESCAQGVGGRVSRVCTSHTGQASHECTEALRPPTTQOCEAK
CDSPTGDPGPECKDNKVAVCPLVLKFCQFSRAYFRMOCCKTCQGH"
BASE COUNT      646 a 1117 c 1083 g 554 t
ORIGIN

```

```

alignment_scores:
  Quality: 791.50      Length: 953
  Ratio: 1.666         Gaps: 43
  Percent Similarity: 49.843  Percent Identity: 28.122

```

alignment\_block:

```

US-10-041-770-2 x AF163762
Align seg 1/1 to: AF163762 from: 1 to: 3400

4 TrpThrGlyArgProTrpLeuLeuLeuLeuLeuSerLeuProgl 20
|||||
526 TGGAAAGGGCGCCATGG...TGGCTGGAGCTTGAAGCCAGCGCTGC 572

20 nLeuGlyLeuAspGlnGluValLeuSerGlyHisLeuGlnThrProt 37
|||||
573 CAGACCCTGGGGAATGAA.....A 592

37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46
|||||
593 CAGAGCGTGGCCAGCCAGCGCTGAAGCGATCGGTACGCCGAGAGCGCTAC 642

46 lTrpGlyProTrp..... 50
|||||
643 GTGGAGACCTGGTGGTGACAAAGATGATGGTGCCTATCACGGGCG 692

51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlnArg 66
|||||
693 CCGGATGTGGAGCAGTATGCTTGGCCATCATCAACATTGTTCCCAAAC 742

67 ArgSerArgThr..... 70
|||||
743 TTTTCCAGGACTCGAGTCTGGGAAGCACCGTTAAACATCTCGTAACTCGC 792

71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeup 85
|||||
793 CTCATCTCTGCACGGAGGACACGCCCACTCTGGAGATCA..... 832

85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101
|||||
833 ..CCCAACATCGCGGGAAGTCCCTAGACAGCTTCTGTAAGTGGCAGAAAT 880

102 ProArgProGlnThrSerProGlnThrLeuProLeuTyArgThr..... 116
|||||
881 CCATCGTGAACACAGCGGCCCATGGCAATGCCATCCAGAACAGCGTGTG 930

```

```

117 .....GlnSerArgGlyArgGlyGlyP 124
931 GCTAACCATGACACAGCAGCGTCTCATCACGCTATGACATCTGCATCTA 980
124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluLeu 140
981 CAAGAACAACACCTTCGGGCACACTAGCTGGCCCGCGTGGCGGGAATGT 1030
141 ArgAlaAlaArgSerArgLeuArgAspProIleLeuProGlyMetPh 157
1031 GTGACGGCGGAGAAGCT..... 1048
157 eGlyTyArgValProPheAlaLeuProLeuHisArgAsnArgArgH 174
1049 .....GCAGCGCTCAATGAGGACATTTGGCCTGCC 1076
174 isProArgSerProProArgSerGlnLeuLeuSerLeuIleSerSer..... 188
1077 ACAAGCGTTCACCATTTGCCACGATGGCCACGATCGGCATGAAC 1126
189 ...ArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPh 204
1127 ATGACGGCGTGGAAACAGCTGTGGGCGCGTGGTGCAGGACCCAGCCA... 1174
204 eSerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSerV 221
1175 ....AGCTCATGGCTGCCACATTAACCATGAAGACCAACCCATTCGTGTG 1220
221 aHisThrProSerProGlnAlaGluProLeuSerProGluThrAlaGln 237
1221 GTCAT.....CCTGCAACCGTGACT 1240
238 ThrGluValAlaProArgThrArg.ProAlaProLeuArgHisHisProA 254
1241 ACATCACCAGCTTCTAGACTCGGCGCTGGGCTCTGCTGAACAACCGG 1290
254 rgAlaGlnAla.....SerGlyThrGluPro..... 262
1291 CCCCCAGACAGACTTTGTGTACCGACAGTGGCACCGGGCCAGCCTA 1340
263 .....ProSerProThrHisSerLeuGlyGlu..... 271
1341 CGATGCAGATGAGCAATGCCGCTTTCAGCATGGAGTCAAAATCGCTCAGT 1390
272 .....GlyGlyPhePheArgAla.....SerProGlnP 281
1391 GTAATACGGGAGGTTCTGCACGAGCTGTGTGTCTGAGCAGAGCAAC 1440
281 ro.....ArgArgProSerSerGlnGlyTrpAlaSerProGln 293
1441 CGGTGCATCACCAACAGCATCCGCGCGCGGAGGCGACGCTGTGCAGAC 1490
294 ValAlaGlyArgArgPro.....AspProPhePr 303
1491 GCACACCATCGACAAAGGGTGTGTACAAACGGGTCTGTGTCCCTTGG 1540
303 oSerValProArgGlyArgGlyGlnGlnGlyProTrpGlyThrG 320
1541 GFTCGCGCCAGAGGTGTGGAGG.....AGCCTGG 1572
320 lyGlyThrProHisGlyProArgLeuLeuProAspProGlnHisProGly 336
1573 GGGCGCTGGACTCCATGGGCGGACTGCAGCGGAC..... 1607
337 AlaTrpLeu.ProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrps 353
1608 CTGTGGCGGGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1644
353 erLeuPheAlaProSerSerProIleProArg.....Cys 364
1645 .....GACAGCCCGAGGCCAACCATCGGGGGCAAGTACTGT 1680
365 SerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProPr 381

```

```

||||| : : : : : ||||| : : : : : |||||
1681 CTGGGTGAGAGAGCGGCGACCGCTCTGCAACAGGATGACTGTCCGCC 1730
381 oGluInProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnG 398
: : : : : ||| : : : : : ||||| : : : : : |||||
1731 TGCTCCAGGAGCTTCAGAGAAAGTGCAGTGTCTGAATTTGACAGATCC 1780
398 luPheMetGlyGlnLeuTyrGlnTyrGluProPheThrGluValGlnGly 414
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
1781 CTTTCGCTGGAAATCTCAAGTGGAAACGATAC...CGGGAGGGGCG 1827
415 SerGlnArgCysGluLeuAsnCysAlaGProArgGlyPheArgPheTyrVa 431
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
1828 GTAAGACGCTGCTCGCTCAGGAGCTAGCGAAGGCTTCAACTTCTACAC 1877
431 larGHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyAlaP 448
1878 GGAGAGGGCGGCGAGCGGTGTGTGACAGGACACCTGCCGCCACGAG 1927
448 roAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGlyIle 464
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
1928 TGGACATTTGCGTCAGTGGCGAATGCAAGCACGTCGGCTGCCAGCGATC 1977
465 LeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyAspAs 481
||||| : : : : : ||| : : : : : ||||| : : : : : |||||
1978 CTGGCTCCGACCTGCGGGAGGACAGTGCAGTGTGTGGCGGTGACGG 2027
481 pSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProL 498
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2028 CAGTGCCTGGAGACCATTCAGGGCGCTTCAGCCAGCCTCAGCTGGGG 2077
498 euGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGln 514
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2078 CCGGTACGAGGATGCTGTGGATTCCCAAGGCTCCGTCCACATCTTC 2127
515 IleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProG 531
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2128 ATCCAGGATCGAACCTCTCTCAGTCACATGTGCCCTGAAGGAGACCA 2177
531 yGlyArgSerIleIleAsnGlyAsnTrpAlaValAspProGlySert 548
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2178 GGAGTCCCTGCTGTGGAGGGCTGCTGGGACCCGCCAGCCCGCACCGTC 2227
548 yrArgAlaGlyThrValPheArgTyrAsnArgProProArgGluGlu 564
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2228 TGCCTTAGCTGGGACCACTTCTCAATGTGCACAGGGGCCAGACCGATC 2277
565 GlyLysGlyGluSerLeuSerAlaGluGlyProThrThrGlnProValas 581
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2278 .....CAGAGCTCGAAGCCCTGGGACCGATTAAATGATCTCTCAT 2318
581 pValTyrMetIlePheGlnGluGluAsnProGlyValPheTyrGlnTyrV 598
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2319 CGTCAATGTGTGGCCGGACCGAGCTGCTGGCCCTCGCTACCGCTTC. 2367
598 alIleSerSer.ProProProIleLeuGluAsnProThrProGluPro.. 613
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2368 .....AATGCCCATCGCCGTGATGCTGCTGCCCC...CCTACTCCTGG 2409
614 .....ProValProGlnLeuGlnProGlu..... 621
2410 CACTATGGCCCTGGACCAAGTGTGCGCCAGATGTGCGAGCGGTAGCCA 2459
622 .....Ile.LeuArgValGluPr 627
2460 GTGCGAGGGTGGAGTCCGCAACCACTGGACAGCTCCGCGGTGCC 2509
627 oProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnA 644
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2510 CCCACTACTGCACTGCCACAGCAAGCT.....GCCCAA 2544
644 rgGlnValArgIleProGln.MetProAlaProProHisProArgThrPr 660
: : : : : ||||| : : : : : ||||| : : : : : |||||

```

```

2545 AGGCAGCGCGCTGCAACACGAGCGCTTGCCTCCAGAC..... 2583
660 oLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysS 677
2584 .....TGGTGTAGGAACCTGGCTCGCTCTGCA 2611
677 exAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSer 693
2612 GCCCAGCTGGATGCAGCGCTGCGACGTGCTGCTGCTGTCACGCGC 2661
694 ArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAlaAl 707
2662 CCGCTCTCTGCGCGGAGAGAGCGCTGACGACAGCAGCGATGCCCGCA 2711
707 aglyAlaArgProProAlaSerProGluProCysHisGlyThrProCysP 724
2712 G...CCGCGCCACCTGTACTG...GAGCCTGCCAGCGCCCACTGCC 2755
724 roProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGly 740
2756 CTCCGAGTGGCGGCCCTCGACTGCTGTAGTGCACCCCGCAGCTGCCGG 2805
741 ProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGly 757
2806 CCGGCGCTCCGCGCACCGCTGCTCTTCAGAGAC.....GCAGACCA 2849
757 yGlySerSerValProProGluArgCysGlyHisLeuProArgProAsnI 774
2850 CCGCGCAGCTGCCCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2899
774 leThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGluVal 788
2900 CCACCATGCTGCTCAACTTGCCTGCTGCTGCTGCTGCTGCTGCTGCT 2949
789 GlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSe 805
2950 GCGCAG...TGGGTGAGTGTCTGCACAGTGGCGCTGCGCGCAGCGCA 2996
805 rArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnG 822
2997 GCGCTCGGTGCGCTGCACGACGACACGCGGCAG.....GCGTCGACG 3040
822 luCysAlaSerGlyProProGlnProProSerArgGluAlaCysasp 837
3041 AGTGCAGGAGGCC...CTGCGCGCCGCCACCGACGAGCAGTGTGTAG 3084

seq_name: gb_pr:AK023772

seq_documentation_block:
LOCUS AK023772 2964 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens CDNA FLJ13710 fis, clone PLACE2000373, weakly similar
to F-SPONDIN PRECURSOR.
ACCESSION AK023772
VERSION AK023772.1 GI:10435805
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE2
clone:PLACE2000373.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fuji,A., Hara,H.,
Tanase,T., Nomura,Y., Togiyasu,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahara,Y., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE
AUTHORS 2 (bases 1 to 2964)
Isogai,T. and Otsuki,T.
Direct Submission
TITLE Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

```





AUTHORS Kapeller-Liebermann, R., Cook W.J. and Silos-Santiago, I.  
TITLE 33428, a human metalloprotease family member and uses thereof  
JOURNAL Patent: WO 0188156-A 10 22-NOV-2001;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES Location/Qualifiers  
source 1. 3132  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 582 a 1043 c 983 g 524 t  
ORIGIN

alignment\_scores:  
Quality: 784.00 Length: 940  
Ratio: 1.701 Gaps: 44  
Percent Similarity: 49.043 Percent Identity: 28.617  
alignment\_block:  
US-10-041-770-2 x AX327754 ..  
Align seg 1/1 to: AX327754 from: 1 to: 3132  
4 TrpThrGlyArgProTrpLeuTyLeuLeuLeuLeuLeuSerLeuProcl 20  
604 TGAAGAGGGGGCCATGG...TGGCTGGGACCTTGAAGCCACCGCTGC 650  
20 nLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuGlnThrProt 37  
651 CAGGCCCTGGGCAATGAA.....A 670  
37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46  
671 CAGAGCGTGGCCAGCCAGGCGCTGAAGCGATCGGTACGCGAGAGCGCTAC 720  
46 lTrpGlyProTrp..... 50  
721 GTGAGACCTCGTGGTGGTGCACAGATGATGGTGCCTATCACGGGG 770  
51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66  
771 CCGGGATGTGGAGCAGTATGCTCTGGCCATCATGAACATTGTTGCCAACC 820  
67 ArgSerArgThr..... 70  
821 TTTTCCAGGACTCGAGTCTGGGAGCACCCTTAACATCTCGTAATCGC 870  
71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85  
871 CTCATCTCTGTCACGGAGGACCAGCCACCTCTGGAGATCA..... 910  
85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
911 ..CCACCATGCCGGGNAAGTCCCTGGACACCTCTCTAAGTGGCAGAAAT 958  
102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThr..... 116  
959 CCATCGTGACACAGCGGCCATGGCAATGCCATTCAGAGAACGGTGTG 1008  
117 .....GlnSerArgGlyArgGlyGlyP 124  
1009 GCTAACCATGACACAGCAGTCTCATCACAGCTATGACATCTGCATCTA 1058  
124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluIle 140  
1059 CAGAACCAACCTCTGGGCACATAGGCTGGCCCGGGTGGCGGATGT 1108  
141 ArgAlaAlaArg.....SerArgLeuArgAspProTleLysProgl 155  
1109 GTGAGCGGAGAGAGTGCAGCGTCAATGAGGACATTTGGCTGGCCACA 1158  
155 yMetPheGlyTyArgValProPheAlaLeuProLeu.HisArgAsn 171  
1159 G.....CGTTCCACCATTTGCCACGAGAT 1181

172 ArgArgHisProArgSerProProArgSerGluLeuSerLeuIleSerSe 188  
1182 CGGCACATTCGCGCATGAACCA.....TGACG 1210  
188 rArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPheS 205  
1211 CGGTGGGAAACAGCTGTGG.....GGCCCTGGTTCAGGACCCAGCC 1251  
205 erAlaAsnGlySerProGln...ThrGluLeuProProThrGluLeuSer 220  
1252 AAGCTCATGGTGGCCCATTAACCATTAAGACCAACCATTCGTGTGTGTC 1301  
221 ValHisThrPro.....SerProGlnAlaGluProLeuSerProGluTh 235  
1302 ATCTGTCAGCCGTGACTACATCACAGCTTTCTAGACTCGGCGCTGGGCG 1351  
235 rAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArgHisH 252  
1352 TCTGCTGAACAACCGGCCCCACAGACAGACTTTGT..... 1388  
252 isProArgAlaGlnAlaSerGlyThrGluPro..... 262  
1389 .....GTACCCGACAGTGGCCCGGCCCAAGCCTACGATGCAGATGA 1430  
263 ProSerProThrHisSerLeuGlyGlu.....G 272  
1431 GCAATGCGCGCTTTCAGCATGGAGTCAATCGCTCAGTGTAAATACGGGG 1480  
272 yGlyPhePheArgAla.....SerProGlnPro..... 281  
1481 AGGTCTGACGAGCTGTGGTGTCTGACCAAGAGCACCCTGTCATCACC 1530  
282 ..ArgArgProSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArg 297  
1531 AACAGCATCCCGCGCGCGAGGCGACGTGTGCCAGACACACCATCGA 1580  
298 ArgPro.....AspProPheProSerValProAr 307  
1581 CAAGGGGTGGTGTACAAACGGGTGTGTCCCTTGGTGTGCGCCCGAG 1630  
307 gglyArgGlyGlnGlnGlyGlnGlyProTrpGlyThrGlyGlyThrProH 324  
1631 AGGGTGTGACGG.....AGCCTGGGGCGGTGGACT 1662  
324 isGlyProArgLeuGluProAspProGlnHisProGlyAlaTrpLeuPro 340  
1663 CCATGGGCGACTGCAGCCGGAC..... 1685  
341 LeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSer.....Le 354  
1686 .....CTGTGGCGCGCGGTGTCT 1705  
354 uPheAlaProSer.SerProIleProArg..... 363  
1706 CTTCTAGCGCTCCTCGACAGCCCGCCAGCCCAACCATCGGGGCAAGTAC 1755  
364 CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380  
1756 TCTCTGGGTGAGAGAGGCGGCGCCCTCTCTGCAACACGGATGACTGTCC 1805  
380 oProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397  
1806 CCTGTGCTCCAGGACTTCAGAGAAGTGCAGTGTCTGAATTTGACAGCA 1855  
397 InGluPheMetGlyGlnLeuTyGlnTrpGluProPheThrGluValGln 413  
1856 TCCCTTTTCGTGGGAAATTTCTAAGTGGAAACCGTAC...CGGGGAGG 1902  
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTy 430  
1903 GCGGTGAGGCGCTGCTCGCTCAGTGGCTAGCGGAGGCTTCAACTTCTA 1952

```
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1953 CAGGGAGGGCGGAGCGGTGGTGGACGGGACCCCTGGCTCCCGCCAGACA 2002
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463
   ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2003 CGGTGGACATTTGGCTCAGTCAGTCGCAATGCAAGCACGTGGCTGCACCGA 2052
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAs 480
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2053 GTCTGGCTCCGACCTGCGGGAGGAGCAAGTCCGAGTGTGGCGGTGA 2102
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2103 CGGCAGTGGCTGGAGACCATCGAGGGCGCTCTCAGCCGAGCCTCACCTG 2152
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 roLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeu 513
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2153 GGGCCGGGTACGAGGATGCTGCTGGATTTCCCAAGGCTCCGTCACATC 2202
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 GlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyPr 530
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2203 TTCATCCAGATCTGAACCTCTCTCTCAGTCACTTGGCCCTGAAGGGAGA 2252
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
530 oGlyGlyArgSerIleAsnGlyAsnTrpAlaValAspProGlyLys 547
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2253 CCAGGAGTCCCTGCTGCTGGAGGGGTGCTGGACCCGCCAGCCACC 2302
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
547 erTyrArgAlaGlyGlyThrValPheArgTyrAsnArgProArgGlu 563
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2303 GTCGTGCTCTAGCTGGGACCACTTTCACTGGCAGCGGGCCAGACAG 2352
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 GluGlyLysGlySerLeuSerAlaGlyGlyProThrGlnProVa 580
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2353 GTC.....CAGAGCTCGAAGCCCTGGGACCGGATTAAATGCATCTCT 2393
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 lAspValTyrMetIlePheGlnGluAsnProGlyValPheTyrGlnT 597
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2394 CATCGTCATGTCGTGCGCGGACCGAGCTGCTGCCCTCCGCTACCGCT 2443
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
597 yrValIleSerSerProProIleLeuGluAsnProThrProGluPr 613
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2444 TC.....AATGCCCCCATCGCCCGTGACTGCTGCCCC...CCTACTCC 2484
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 o.....ProValProGlnLeuGlnProGlu..... 621
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2485 TGGCACTATGCGCCCTGGACCAAGTCTGGCCCAAGTGTGACGGCGGTAG 2534
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 .....IleLeuArgValG1 626
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2535 CCAGGTCAGCGCGTGGAGTGGCGCAACAGCTGGACAGCTCCGCGGTG 2584
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2585 CCGCCCACTACTGCAGTGGCCACAGCAAGCT.....GCC 2619
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
643 lnArgGlnValArgIleProGlnMetProAlaProProHisProArgTh 659
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2620 AAAGCAGCGCGCTGCAACAGCGAGCCTTGCCCTCCAGAC..... 2661
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 rProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaC 676
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2662 .....TGGGTTGTAGGGAACCTGCTGCTCT 2686
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 ysSerAlaSerCysGlyValTrpArgProIlePheLeuCysIle 692
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2687 GCAGCCGACGCTCGATGCGGCGGAGCGGCGCGCTCGGTGCTGTCAG 2736
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
693 SerArgGluSerGlyGluGlu.....LeuAspLysArgSerCysAl 706
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2737 CGCGCGCTCTGCGCGGAGAGAGGCGCTGGACGACAGCGCATGCC 2786
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
706 alaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProC 723
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 13:01:30 ; Search time 2031.65 Seconds  
(without alignments)  
17498.575 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 agggagaactggactggcag.....cctcagccttcacgcatag 2634

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	652.6	24.8	891	9	AL513944
2	497.4	18.9	744	10	BF536373
3	449	17.0	597	9	AI683106
4	419.6	15.9	865	10	BI094238
5	411	15.6	437	10	BM148104
6	391.4	14.9	517	10	BE808908
7	373.4	14.2	495	9	AW230357
8	370.8	14.1	625	10	BG438243
9	359.8	13.7	866	10	BI330332
10	357.4	13.6	360	10	BF993048
11	338	12.8	592	10	BM391602
12	337	12.8	370	10	BG005397
13	330.8	12.6	513	9	AW763191
14	317.8	12.1	375	10	BG008790
15	314.4	11.9	324	10	BF993027
16	305.2	11.6	449	10	BG900891
17	303.8	11.5	373	10	BF987707

## ALIGNMENTS

RESULT 1

AL513944

LOCUS AL513944 LTI\_NFL006\_PL2 891 bp mRNA linear EST 13-FEB-2001

DEFINITION prime, mRNA sequence.

VERSION AL513944.1 GI:12777438

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 891)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Source 1..891

Location/Qualifiers

1..891

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CLOBA010ZE04"

/clone\_lib="LTI\_NFL006\_PL2"

/tissue\_type="placenta"

/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(gt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 162 a 305 c 265 g 155 t 4 others

ORIGIN

BF989747 IL5-GN017  
AI637480 tt07h07.x  
BF993841 IL5-GN017  
D78761 HUM510H03B  
BF986444 QV4-GN014  
BE757541 212006 MA  
BF989357 IL5-GN017  
BM150365 TCAAP9Q92  
BM855583 BB855583  
BM149824 TCAAP3E92  
BF997426 PM1-GN018  
BI836313 603083995  
BG008820 MR3-GN018  
BF870887 IL0-ET015  
BG92879 342468 BA  
BF367876 KC4-GN003  
AI521594 to65c12.x  
BF997840 MR2-GN012  
BF997839 MR2-GN012  
BI645713 603275026  
BF443732 261436 MA  
BF651524 274324 MA  
BB862174 BB862174  
BF412563 UI-R-BT1-  
BB857367 BB857367  
BB856356 BB856356  
AZ335574 LM0065P2  
BB643815 BB643815

287.2 10.9 307 10 BF989747  
282 10.7 309 9 AI637480  
279 10.6 327 10 BF993841  
268.2 10.2 286 10 D78761  
261.4 9.9 507 10 BF986444  
260.8 9.9 571 10 BE757541  
260.4 9.9 262 10 BF989357  
239.2 9.1 266 10 BM150365  
221.4 8.4 498 9 BB855583  
219 8.3 274 10 BM149824  
214.4 8.1 273 10 BF997426  
205.6 7.8 711 10 BI836313  
195.2 7.4 389 10 BG008820  
193 7.3 393 10 BF870887  
183.2 7.0 284 10 BG92879  
182.2 6.9 216 10 BF367876  
181.6 6.9 380 9 AI521594  
176 6.7 214 10 BF997840  
170.8 6.5 183 10 BF997839  
159.8 6.1 775 10 BI645713  
153.6 5.8 479 10 BF443732  
146.4 5.6 480 10 BF651524  
146.2 5.6 446 9 BB862174  
140 5.3 510 10 BF412563  
139.8 5.3 379 9 BB857367  
137 5.2 483 9 BB856356  
134 5.1 664 12 AZ335574  
134 5.1 691 9 BB643815

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9519 row: g column: 08  
High quality sequence stop: 685.

FEATURES  
source  
1. .744  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4191007"  
/lab\_host="NCI\_CGAP\_SG2"  
/lab\_host="DH10B (T1 phage-resistant)"  
Note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
Note1: Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 139 a 231 c 208 g 166 t

Query Match 18.9%; Score 497.4; DB 10; Length 744;  
Best Local Similarity 79.3%; Pred. No. 3.5e-92;  
Matches 618; Conservative 0; Mismatches 126; Indels 35; Gaps 1;

QY 1281 ccgcttctatgtccgtcacactgaaagggtccaggatggagccctgtgtcagcctggagc 1340  
Db 1 CCGATTCTATGTCCGACACACTGAAAGGTCAGGATGGAACCTGTGTGAGCTGGATC 60  
QY 1341 ccttgacatctgtgtggtgagcgtgtctgagcccgctgtgaggtggtccttggctc 1400  
Db 61 CCTAGACATCTGTGTGGCTGGACGCTGCCCT----- 90  
QY 1401 tggcaggcgtccctgatgtgtgagctgtgtgggtgtgattctacctgtgccttgt 1460  
Db 91 -----GCGCTCCGATGGCTGCGAGTCTCGGGGGTGGTGTCTACCTGCCGCTGGT 145  
QY 1461 ttccgggaacctcaactgacccaggggggggggggggggggggggggggggggggggg 1520  
Db 146 TTCGGGAAATCTCACTACCCAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 205  
QY 1521 agcgggagccttgcggctccagattgcccagctccggcctagctcccaactacctggca 1580  
Db 206 TCGCGGGGCTCCACCTTCACATTTCCAGTTGCGAGCCAGTCCCAATACCTCGCACT 265  
QY 1581 tctgtggccctgggggggggggggggggggggggggggggggggggggggggggggg 1640  
Db 266 CCGAGGGCTTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 325  
QY 1641 ctacaggcggcgggggggggggggggggggggggggggggggggggggggggggggg 1700  
Db 326 CTATACAGCCATTGGGACTGCTCTCCAGTATACCGTCTCCACGGGAGAGGAGGG 385  
QY 1701 ggagagtctgtcgggtgaaggcccccagcccccagcccccagcccccagcccccagcc 1760  
Db 386 GGAGAGTCTGTACAGAGAAGGCCCTACCACTAGCCCTGTGGAGCTGTATATGATCTTCA 445  
QY 1761 ggaggaacacccagggcgttttttatcagtagtgcattctctcaccctccctcctga 1820  
Db 446 AGAGGACACCCAGGTGTTTTTATCATGATGATGATGATGATGATGATGATGATGAT 505  
QY 1821 gaacccacccagagagcccccgtgtcccccagcccccagcccccagcccccagccccc 1880  
Db 506 GAGTCTTCCACAAAAGCCCTCAGGCCCTTCAGCCCTTCAGCCCTGAGATGCTGAGGGGG 565  
QY 1881 cccactgtccggcaccggccggccggccggccggccggccggccggccggccggccggcc 1940  
Db 566 CCTACTCCCTCAGCCCCCGCCGAGTCCGGGACCCAGGACCCCTCCCACTCAGGTGG 625  
QY 1941 gatcccccagatgcccgggggggggggggggggggggggggggggggggggggggg 2000

Query Match 24.8%; Score 652.6; DB 9; Length 891;  
Best Local Similarity 99.0%; Pred. No. 4.4e-124;  
Matches 664; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 1 atggagaactgactggcagggccctggctgtatctgtctgtctgtctgtctgtctgtct 60  
Db 222 ATGGAGAACTGACATGGAGGCCCTGGCTGTATCTGTCTGTCTGTCTGTCTGTCTGTCT 281  
QY 61 ctctcttgatcaggaggtgtgtccggagacactctctcagacacctacagagagggc 120  
Db 282 CTTCTGCTGGATCAGAGAGGTGTGTCCGACACACTCTTTCAGACACTCAGAGAGGGC 341  
QY 121 caggggcccccgaagtgtctgggggaacttgggtccagtgaggccttctgtcccccctgc 180  
Db 342 CAGGGCCCGGAA-GTGTCTGGGAGCTTGGGTCCAGTGGGCTTGTGTCCAGCCCTGC 400  
QY 181 ggggtgggggtgacgcagagccggacatgtagctccctacagtcagctcacccg 240  
Db 401 GGGGTGGGGTGCACGAGGAGCCGGACATGTACCTTCCCTCAGTCCAGTCCAGTCC 460  
QY 241 agtctgcccctccctcccccgggggggggggggggggggggggggggggggggggggg 300  
Db 461 AGTCTKCHCTCCCTCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 520  
QY 301 ggtccagagccagacactctccagaaacccctccctgttacagagacagctctcggga 360  
Db 521 GTTCCAGACCCAGACTTCTCCAGAAACCCCTCCCTTGTGTACGAGACAGTCTCGGGGA 580  
QY 361 aggggtggcccacttcgaggtcccttcccacactagggagagagagacccagagatt 420  
Db 581 AGGGGTGGCCACTTCGAGGTCCGCTTCCACCTTAGGAGAGAGAGAGACCCAGAGATT 640  
QY 421 cagagggccaggaggtccgggttcagagcccccagcccccagcccccaggaatttgcgttat 480  
Db 641 CGAGGGCCAGAGTCCCGGCTTCAGACCCCATCAAGCCAGGAGGATGTCGGTATGGG 700  
QY 481 agagtgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgcct 540  
Db 701 AGAGTGCCCTTTGCATTGCCACTGCACCGGACCCGACACACCTCGAGGCCACCCAGA 760  
QY 541 ttgtagctgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 600  
Db 761 TGTGAGTCTCCCTGATCTCTTCTAGAGGGGAGAGCCATATTCGCTCCCTCTACTCAA 820  
QY 601 gcagagcaattctccgaaacggagcccccacactgagctccctcccccagagactgtct 660  
Db 821 GCAGAGCCATTCTCCGAAACGGCAGCCGCCCAACTGAGTCCCTCCACAGAGTGTCT 880  
QY 661 gtccacacccc 671  
Db 881 GTCCACWCCC 891

RESULT 2  
BF536373 744 bp mRNA linear EST 11-DEC-2000  
LOCUS 602051728F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4191007 5',  
DEFINITION mRNA sequence.  
ACCESSION BF536373  
VERSION BF536373.1 GI:11623741  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 744)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.

Db 626 GATCCCAAGTCCCTCCGACACTGCTGTCAGGACACCAAGGGTCTTCTGCTGGATA 685

QY 2001 ctggaaacagtgaggacactctgcagtcgctcgcgggaaagctgtctgcgcgc 2059

Db 686 CTGGGAAACAGTGGGGCACTCTGAGTGTTCAGCATCCTGTGGGCAAGGTGTTTGACCC 744

RESULT 3

AI683106/c

LOCUS

DEFINITION

AI683106 597 bp mRNA linear EST 16-DEC-1999

similar to X01411.x1 NCI-CGAP\_Ut4 Homo sapiens cDNA clone IMAGE:2267925 3',

repetitive element ;, mRNA sequence.

ACCESSION

AI683106

VERSION

AI683106.1 GI:4893278

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 597)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: Greg Lennon, Ph.D.

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 1132 Std Error: 0.00

Seq Primer: 40UP from Gibco

High quality sequence stop: 369.

Location/Qualifiers

1..597

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2267925"

/clone\_lib="NCI-CGAP\_Ut4"

/tissue\_type="serous papillary carcinoma, high grade, 2

pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.48 kb. Life Technologies catalog #:

11542-016"

BASE COUNT 118 a 162 c 180 g 136 t 1 others

ORIGIN

Query Match 17.0%; Score 449; DB 9; Length 597;

Best Local Similarity 89.6%; Pred. No. 3e-82;

Mismatches 0; Gaps 0;

Matches 482; Conservative 0; Indels 56; Indels 0; Gaps 0;

QY 133 ggtgtctgggacattgggtccagtgaggctcttcttcacagccctcgagggtgggtg 192

Db 597 GGTGTTTGGGGCCCTTGGGTCAAGTGGGCTCTTGTCTCAAGCCCTGCGGGGTGGGTG 538

QY 193 cagcgagagccggagacatgcagtcctcctacagtcagctccaccagctgcctc 252

Db 537 CAGCGAGAGCGGCAACATGTCATCTCTTACAGTGGAGCTCTCACCGGAGTTGGCCCTC 478

QY 253 cctccccggcccccaagacatccagaagccctcctccccggggccaggggtcccaagacc 312

Db 477 CTCTCCGGGCCCCCAAGACATCCAGAAGCTTCCTCCCCGGGGCCAGGGTCCACAGACC 418

QY 313 cagacttctccagaaccctcccttctgtacaggacacagctctcggggaaggggtggccca 372

Db 417 CAGATTCTCCAGAAACCTCCCTTTGTACAGACCCAGTTTCGGGAGAGGGTGGCCCC 358

QY 373 cttogaggtcccgcttccacacctagggagagagagagagagagagagagagagagag 432

Db 357 TTTCGGGGTCCCGTTCCCGTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298

QY 433 aggtcccggttcagagaccccatcaacaggaatgttcggttatggagagagagagagag 492

Db 297 GGTGTCCTGGTTCGAGGCCCTCCATCAAGCCAGGATGTGGGTATGGGAGAGTGCCTTT 238

QY 493 gcattgcactgcacccggaacccagcagccctcggagcccccacccagatctgagctgtcc 552

Db 237 GCATTGCCANTGCAGGGGAAACCGGGGCCCTGGGGGCCCTCCAGTTTGTAGCTGTCC 178

QY 553 ctgactctctagaggggaagaggtcttcctccctactccactccacagagagagagagag 612

Db 177 CTGTTTTCCTAGGGGAGAGGGGTTTTCGGTCTCCCTACTCCAGAGAGAGAGCTTTT 118

QY 613 tccgaacagcgagcccccaaacatgagctccctcccccacagaactgtctgtccacacc 670

Db 117 TCCGAAAGGGCGGCCCCCAAAATTTGGGCTCCCTCCCGCAGAACTGTTTTCCTCCGCC 60

RESULT 4

BI094238

LOCUS

DEFINITION

602860118F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:5001173 5',

mRNA sequence.

ACCESSION

BI094238

VERSION

BI094238.1 GI:14512568

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 865)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L14M1034 row: d column: 06

High quality sequence stop: 510.

Location/Qualifiers

1..865

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5001173"

/clone\_lib="NIH\_MGC\_10"

/cell\_line="MGC36"

/lab\_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.5 kb. Library prepared by Life

Technologies.

BASE COUNT 240 a 187 c 303 g 135 t

ORIGIN

Query Match 15.9%; Score 419.6; DB 10; Length 865;

Best Local Similarity 92.3%; Pred. No. 3.8e-76;

Mismatches 0; Mismatches 34; Indels 9; Gaps 7;

Matches 519; Conservative 0; Indels 9; Gaps 7;

QY 1273 cgtggcttccgcttctatgtccgtcacactgaaaggtccagagagagagagagagag 1332

Db 1 CGTGGCTTCGGCTTATGTCCGTACACTGAAAGGTCCAGGATGGAGCCCTGTGTAG 60  
QY 1333 cctggagccctgacatctgtgtgctgagcgtctgtgagcccccgtgtgatggatc 1392  
Db 61 GCTGGAGCCCTGACATCTGTGGCTGGACGCTGTCTGAGCCCGCGCTGTGATGGGATC 120  
QY 1393 attggctctggcagcgctcctgatagc-tgtgagctgtgtgggggtgatattacatg 1451  
Db 121 CTGTGCTCTGGCAGCGCTCTGTGATGGCTGTGGAGTCTGTGGGGGTGATGATCTACCTG 180  
QY 1452 tgcctcttctggggaacctcaactgacccggggggcccccctggctgctatcagaagatctt 1511  
Db 181 TCGCCTTCTTCGGGGAACCTCACTGACCGAGGGGGCCCTGGCTGATCAGAGATCTT 240  
QY 1512 gtgattccagc-gggagccttgaggctccagattgcccagctccggcctagctccaaact 1570  
Db 241 GTGGATTCAGCGGGGAGCCCTTGGGCTGCCAGATG-CCAGCTCCGGCCTAGCTCCAACT 299  
QY 1571 acctggcaactctgtgacctgggggccc-ggtccatcatcaatgggaac---tgggctgtg 1626  
Db 300 ACCTGGCACTTCGTGGCCCTGGGGCCCTGGTCCATCATCAATGGGAATCTGTGGCTGTG 359  
QY 1627 gatccccctgggtctacagcggcggggagccgtcttcttcgataataaccgtcccccagg 1686  
Db 360 GATCCCTCTGGTCTACAGGCGCGGGACCGCTTCTTCGATATATAACCGTCCCGAGG 419  
QY 1687 g-aggagggcaaggggagctgtcgtgctgaagggcccccaccagcctgtggatgt 1745  
Db 420 GAAGGAGGCAAGGGAGAGTCTGCGCTGAGG-CCCAACCAACAGCCCTGTGGATGT 478  
QY 1746 ctatatgatcttcaggaggaaaccccggtttttatcagtatgtcatctcttccacc 1805  
Db 479 CTATATGATCTTTCAGGAGGAAACCCAGGGGTAGATAATCGAGTAGTGTCTCGAGTCTCT 538  
QY 1806 tctccaatccttgagaacccc 1827  
Db 539 GTACAGCGTCTGCTCGCAATACC 560

RESULT 5  
BM148104  
LOCUS  
DEFINITION  
TCAAP108352 Pediatric acute myelogenous leukemia cell (FAB M1)  
Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP8352, mRNA  
sequence.  
ACCESSION  
BM148104  
VERSION  
BM148104.1 GI:17168519  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1 (bases 1 to 437)  
AUTHORS  
Wei,Y., Teang,Y.T.M., Mei,S., Ku,J.M., Ali-Ösman,F.R. Jr.,  
Guanaratne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.  
Pediatric Leukemia cDNA Sequencing Project (2001)  
Unpublished (2001)  
COMMENT  
Contact: Dr. Judith F. Margolin  
Texas Children's Cancer Center and Human Genome Sequencing Center  
at Baylor College of Medicine  
1102 Bates, MC3-3320 Houston, TX 77030, USA  
Tel: 832-824-4536  
Fax: 832-825-4038  
Email: clones@txccc.org  
Seq primer: M13 primer.  
Location/Qualifiers  
1..437  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TCAAP8352"  
/clone\_lib="Pediatric acute myelogenous leukemia cell (FAB  
M1) Baylor-HGSC project-TCAA"  
/sex="male"

/issue\_type="leukopheresis"  
/cell\_type="myeloid cell"  
/dev\_stage="pediatric 6 years"  
/lab\_host="DH10B"  
/note="Vector: lambda pSB: Site 1: BamHI; Site 2: EcoRI;  
First strand cDNA was primed with an anchored  
XhoI-oligo(dT) primer [5'GGAGACTCGAGGGCGCCGAGGAGAG(T)YN  
3'; V-A,C,G; N-A,C,G,T] and then dg tailed. Second strand  
was primed with a BamHI-dC primer  
[5'AGAGAGCTCGATCCGGCGCCCAATAATAAT(C) 3'].  
Double-stranded cDNA was then digested with BamHI and XhoI  
and directionally cloned into the BamHI and SalI sites of  
lambda pSB vector. Library went through one round of  
normalization. Library was constructed by Wei Yu at RIKEN  
of Japan (Carninci P, Westover A, Nishiyama I, Ohsumi T,  
Itoh M, Nagao S, Sasaki N, Okazaki Y, Muramatsu M,  
Schneider C, Hayashizaki Y, High efficiency selection of  
full-length cDNA by improved biotinylated cap trapper.,  
DNA Res 4: 1, 61-6, Feb 28, 1997).  
BASE COUNT 76 a 151 c 131 g  
ORIGIN

Query Match 15.6%; Score 411; DB 10; Length 437;  
Best Local Similarity 97.7%; Pred. No. 1.7e-74;  
Matches 417; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 tggactggcagggccctggtatctgctgctgtctctctccctccctcagctctcttg 69  
Db 7 TTGTTTGGAGGCCCTGGCTGTATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 66  
QY 70 gatcaggaggtgtgttcgagacactctcttcagacacacacacagagagggccagggcccc 129  
Db 67 GATCAGGAGGTGTTGCCGGACACTCTCTTCAGACACCTACAGAGGAGGGCCAGGGCCCC 126  
QY 130 gaagggtgtctggggaccttggtccagtggtggtcctcttctccagccctcggggtggggg 189  
Db 127 GAAGGTGTCTGGGGACCTTGGGTCCAGTGGGCTCTGTCTCCAGCCCTGCGGGGTGGGG 186  
QY 190 gtgcagcgcagggcgggacatgtcagctccctacagtcagctccacccagctcttgcgcc 249  
Db 187 GTGCAGCGCAGGAGCGGACATGTCTAGTCTCTCTACAGTGAGTCTCCACCGAGTCTGCC 246  
QY 250 ctccctcccgcccccacagacatccagagccctctctcccgggggccagggctccaga 309  
Db 247 CTCCTCCCGCGGCCCAAGACATCCAGAGCCCTCTCTCCCGGGGCCAGGGTCCCGAGA 306  
QY 310 cccagactctccagaaacccctccctttgacagacacacagctctcggggaagggtggc 369  
Db 307 CCCAGACTTCTCCAGAAACCCCTCCCTTTGTACAGGACACAGTCTCGGGGAAGGGGTGGC 366  
QY 370 ccacttcgaggtcccgcttccacactaggagagagagagagagagagagagagagagag 429  
Db 367 CCATTCAAGAGCCCGCTTCCACCTAGGAGAGAGAGAGAGAGAGAGATTCAGAGCGGAC 426  
QY 430 agggaggt 436  
Db 427 AGCAGGT 433

RESULT 6  
BE808908  
LOCUS  
DEFINITION  
214215 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION  
BE808908  
VERSION  
BE808908.1 GI:10240020  
KEYWORDS  
EST.  
SOURCE  
cow.  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 517)



**AUTHORS**  
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keeler, J.W.

**TITLE**  
Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

**JOURNAL MEDLINE COMMENT**  
Genome Res. 11 (4), 626-630 (2001)  
Contact: Smith, TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAAACAGCTATGACCAT  
BACKWARD: GTTTCCTCCAGTCACGAG  
Plate: 70 row: N column: 6  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers  
1. .517  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2Bov"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORT6; Site: 1: XhoI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."  
84 a 175 c 179 g 79 t

**BASE COUNT**  
ORIGIN

**FEATURES**  
source  
1. .517  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2Bov"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORT6; Site: 1: XhoI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."  
84 a 175 c 179 g 79 t

**Query Match**  
Best Local Similarity 14.9%; Score 391.4; DB 10; Length 517;  
Matches 428; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 2071 tgcattccgtagtcgagaggaacgagtgatgaacgagctgtgcccgggtgcccagg 2130  
|||||  
Db 1 TGCATTCTCAGAGTCAGAGAGAGAGCTGGATGAACGAGCTGTGCCATGGTGCCAGG 60  
|||||  
QY 2131 cccccagctccctgaacccctgcaacgagccacccatgcccccatctggaagagctgac 2190  
|||||  
Db 61 CCCCCAGCTCCAGAGAGCTGCCAGGCCCCCGCTGCCACCATACCTGAGGAGCCGCG 120  
|||||  
QY 2191 gattgacatctgagccgctctctgcccgcgacccagcaccgagctgagtgac 2250  
|||||  
Db 121 GATGGACGTCCTGCAGCCGCTCTGTGGACCGGACCCAGCAGCCGCTCAGCTACGCTGC 180  
|||||  
QY 2251 cggcaggaatttgggggggtgctcctcgtggtgcccccgagcgtgtggaacatctccc 2310  
|||||  
Db 181 CGCAGAGAGTTTGGGGTGGCGGCTCTCAGTCCGCTAGAGCGCTGCGGACCTGCGC 240  
|||||  
QY 2311 cggcccaacatccagctctccagctgctcgtggtgcccccgagcgtgtggaagtgctct 2370  
|||||  
Db 241 CGACCCCAACATCACCCAGCCCTGCCAGCTGCGCTCTGTGGCCATTTGGAGGTTTCGCTCA 300  
|||||  
QY 2371 ccttggagccagtgctcgtgctgctgctgctgctgctgctgctgctgctgctgctgct 2430  
|||||  
Db 301 CCTGGAGTCAGTGCTCTGTGCGATGCGGCGCGGTCAGAGGAGCGCGGAGGTCCTGT 360  
|||||  
QY 2431 gttgggaacacagtgatgaagtgaagcagcagcagcagcagcagcagcagcagcagcagc 2490  
|||||  
Db 361 GTCGGCAACACCGGGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420  
|||||  
QY 2491 cccagcagagagcctgtgacatggggccctgtactactcctggttccacacgagctgg 2550  
|||||  
Db 421 CCCAGCAGAGAGGCTCGACATGGGGCCCTGACACAGCGCTTGGTTCACACAGGACTGG 480  
|||||

QY 2551 agctccaag 2559  
|||||  
Db 481 AGTCCCAAG 489

**RESULT 7**

**AW230557**  
**LOCUS**  
DEFINITION  
u065all.y1 NCI-CGAP\_Mam1 Mus musculus cDNA clone IMAGE:2647388 5'  
similar to TR:060345 O60345 KIAA0605 PROTEIN. ; mRNA sequence.  
**ACCESSION**  
AW230557  
**VERSION**  
AW230557.1  
**KEYWORDS**  
EST.  
**SOURCE**  
house mouse.  
**ORGANISM**  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE**  
1 (bases 1 to 495)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**AUTHORS**  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**TITLE**  
Tumor Gene Index  
**JOURNAL**  
Unpublished (1997)  
**COMMENT**  
Other ESTs: u065all.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@email.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

MG1:1027840

Seq primer: -40RP from Gibco  
High quality sequence stop: 357.

**FEATURES**  
Location/Qualifiers

1. .495  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2647388"  
/clone\_lib="NCI-CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: PCMV-SPORT6; Site: 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
84 a 155 c 149 g 107 t

**BASE COUNT**  
ORIGIN

**Query Match**  
Best Local Similarity 14.2%; Score 373.4; DB 9; Length 495;  
Matches 419; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1939 cggatccccagatcccgccccccatccaggacacccctgggtctccagctg 1998  
|||||  
Db 1 CGATTCCCAAGTCCCTCCCTCCGACTGCTGTGTCAGACAGCATGGGCTCTCTCTGGA 60  
|||||  
QY 1999 tactggaacagagtgaggacactctgctcagcgtctcgcgggaaagggtctcggcgc 2058  
|||||  
Db 61 TACTGGAACAAGTGGGGCAGCTCTGAGTCTTCAGCATCCTGTGGCAAGGTTTGGCAC 120  
|||||  
QY 2059 cccatttctctgactctccgtagtcggagaggaactggagcagcagctgtgccc 2118  
|||||  
Db 121 CCCATTTCCTCTGATTTCCCGTAGTCCGAGGAGGAGTGGATGACACAGAGCTGTGCT 180  
|||||  
QY 2119 gcgggtgcagggccccccagcctcccctgaacctgcccacggccacccccatgccccccatc 2178  
|||||

```
Query Match      14.1%; Score 370.8; DB 10; Length 625;
Best Local Similarity 98.8%; Pred. No. 3.7e-66;
Matches 405; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
```

```

source
i. .886
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:513860"
/clone_lib="NCI_CGAP_b19"
/lab_host="DH10B (T1 phase-resistant)"
/notes="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
149 a 332 c 237 g 148 t
BASE COUNT

```

	BASE COUNT	ORIGIN
1	0.000	0.000
2	0.000	0.000
3	0.000	0.000
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000
14	0.000	0.000
15	0.000	0.000
16	0.000	0.000
17	0.000	0.000
18	0.000	0.000
19	0.000	0.000
20	0.000	0.000
21	0.000	0.000
22	0.000	0.000
23	0.000	0.000
24	0.000	0.000
25	0.000	0.000
26	0.000	0.000
27	0.000	0.000
28	0.000	0.000
29	0.000	0.000
30	0.000	0.000
31	0.000	0.000
32	0.000	0.000
33	0.000	0.000
34	0.000	0.000
35	0.000	0.000
36	0.000	0.000
37	0.000	0.000
38	0.000	0.000
39	0.000	0.000
40	0.000	0.000
41	0.000	0.000
42	0.000	0.000
43	0.000	0.000
44	0.000	0.000
45	0.000	0.000
46	0.000	0.000
47	0.000	0.000
48	0.000	0.000
49	0.000	0.000
50	0.000	0.000
51	0.000	0.000
52	0.000	0.000
53	0.000	0.000
54	0.000	0.000
55	0.000	0.000
56	0.000	0.000
57	0.000	0.000
58	0.000	0.000
59	0.000	0.000
60	0.000	0.000
61	0.000	0.000
62	0.000	0.000
63	0.000	0.000
64	0.000	0.000
65	0.000	0.000
66	0.000	0.000
67	0.000	0.000
68	0.000	0.000
69	0.000	0.000
70	0.000	0.000
71	0.000	0.000
72	0.000	0.000
73	0.000	0.000
74	0.000	0.000
75	0.000	0.000
76	0.000	0.000
77	0.000	0.000
78	0.000	0.000
79	0.000	0.000
80	0.000	0.000
81	0.000	0.000
82	0.000	0.000
83	0.000	0.000
84	0.000	0.000
85	0.000	0.000
86	0.000	0.000
87	0.000	0.000
88	0.000	0.000
89	0.000	0.000
90	0.000	0.000
91	0.000	0.000
92	0.000	0.000
93	0.000	0.000
94	0.000	0.000
95	0.000	0.000
96	0.000	0.000
97	0.000	0.000
98	0.000	0.000
99	0.000	0.000
100	0.000	0.000

```
Query Match      13.7%; Score 359.8; DB 10; Length 866;
Best Local Similarity 76.6%; Pred. No. 7.6e-64;
Matches 557; Conservative 0; Mismatches 157; Indels 13; Gaps 9;

QY 1125 ccaagcgcctgccccctgagcagcagacccccgggccccctgagtcgagcagccttaa 1184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CCAGGAGCCTTGCCCCCTGAGCATGCCAGACCCAGGACCCCTACAGTGCCTGCGCTTTGA 60

QY 1185 ctcacagaattcatggccagctatcatgagggagcccttcactgaagtcacagggctc 1244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTCCAGGAATTCATGGCCAGCTGTACAGTGGAGGCCCTTCCAGGAAGTTCAGGGGCTC 120

QY 1245 ccaagcgtgtaactgaactgcggccccctggcttcgcgtcttatgtccgtcacactga 1304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CCAGGCGCTGTAACCTGAACCTGCCGCCCGCTGGCTTCGGATTCATGTCGACACACTGA 180

QY 1305 aaagtcagagtgagccctgtgcaacctgagacccctcaactctgtgctggaagc 1364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 AAAGTGCAGGATGAACCCCTGTGACGCTGTATCCCTAGACATCTGTGTGGCTGGAGC 240

QY 1365 ctgctcgaagccgcgtgtgagggatccttggtcctggcagggcgtctg-atggctgtg 1423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CTGCTGTGACCCCTGCTGTGATGGGTCTTGCTGCTGTGCAGGCGTCCGGCATGCTGCG 300

QY 1424 gagctgtgggggtga-tgattctacctgtgccttggtttgggg-gaacctcaactgaagc 1481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GAGTCTGCGGGGGTGAAGTGGTCTACCTGCCCGCTGGTTCGGGCAAAATCTCACTGACCG 360

QY 1482 agggggccccctgggctatcagaa--gatcttggattccagcgggagccttgcggctc 1539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 AGGGGCCCCCTTGGGCTATCAGATCGATCCGTGTGATCCCTGCCGGGCGCTCCACACTT 420

QY 1540 cagattgccagctcggcctgagctcaactacctggaactctgtgcccctggggggcgg 1599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 CACATTTCCAGTTGGGACC-AGTTCCATTTACCTCGCACTCCGAGGGCCTTCGGGGCGCG 479

QY 1600 tcaatcaatgggaactggcgctggtatccccctggg-tccatcagggcgccgcggga 1657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 TCATTCATCACTGGGAATGGGCTGTGATCTCCAGGGCTCCCTACAGCCATCGCGGA 539

QY 1658 cegcttttcgatataaccctctctccacgg-----gagggaggcgaaggagagtgctgtcg 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 CTGCTCTCCAGTATAGCGTCTCCACGGGCAAGACAGAGGCCACCGGACGACGAGTCTGCA 599

QY 1714 gctgaagggcccaaccacagcgtgtg-gatgtctatgatctttcaggaggaagaaacc 1772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 GCCAAAGGCGCTTACCACCTCAGCGCTGGCGCGCGCTCTCTCTGCAACCCACCGAGGACC 659

QY 1773 aggcgtttttatcagatgtatctcttcacactctcctcaatccttgagaaacccccc 1832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 660 CCCAGGCTGTGCGCGACCCCGCATGCTCATCCACCCCTCCCGCGGCGACCCGACAG 719

QY 1833 agagccc 1839
    ||| |||
Db 720 AGGCCCC 726

RESULT 10
BF993048/c 360 bp mRNA linear EST 23-JAN-2001
LOCUS IL5-GN0178-311000-204-f10 GN0178 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF993048
ACCESSION BF993048
VERSION BF993048.1 GI:12399371
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zagoi,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-GN0178-
311000-204-f10&t3=2000-10-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 305.
Location/Qualifiers
source
1..360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0178"
/dev_stage="Adult"
/note="Organ: Placenta_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 59 a 80 c 142 g 78 t 1 others
ORIGIN
Query Match 13.6%; Score 357.4; DB 10; Length 360;
Best Local Similarity 99.4%; Pred. No. 1.7e-63;
Matches 358; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 470 tcggttatggagagtgcccttgccattgccaactgcaacggaaacccagcagccctcgga 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TCGGTTATGGAGAGTGCCCTNTGCATTGCCACTGCCCGGAACCCGAGCACCCTCGGA 301

QY 530 gccacccagatctgagctgcccctgatctcttagagggaagagctattccgtccc 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 GCCCACCAGATCTGAGCTGTCCCTGATCTCTTAGAGGGGAAGAGCCTATTCCGTC 241

QY 590 ctactccaagagcagagccattctccgcaaacggcagcccccaactgagctccctccca 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 CTACTCCAAGAGCAGAGGCATTCCTCCGCAAAACGGCAGGCCCAAACTGAGCTCCCTCCA 181

QY 650 cagaactgtctgcacacccatcccccccagcagacctctaagccctgaactctc 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 CAGAACTGTCTGTCCACACCCCTCCCTCCCAAGCAGAACCTCTTAGCCCTGAACTGCTC 121

QY 710 agacagagtggtgccccagaaacagggcctgccccctacggcatcacccccagagccag 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 AGACAGAGGTGGCCCCCAGAACCCAGGCGCTGCCCTACGGCATCACCCAGAGCCAGG 61

QY 770 cctctggcagagagccccctcaccacagcactcttaggagaagggtggtcttcctg 829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 CTTCTGGCACAGAGCCCCCTTCACCCAGCAGCTCTCTTAGGAGAGGCTGCTTCTTCCGTG 1
```

```
RESULT 11
BM391602 592 bp mRNA linear EST 17-JAN-2002
LOCUS BM391602
DEFINITION UI-R-DY0-ckr-1-07-0-UI.s1 UI-R-DY0 Rattus norvegicus cDNA clone
AUTHORS UI-R-DY0-ckr-1-07-0-UI 3', mRNA sequence.
ACCESSION BM391602
```

QY	2398	ggcgcggggccagagaagccgcgaggttcgtctgtgttgggaacacaggtgtagaagtgagc	2457
Db	331	ggtctgtgttcagagagagccgcaagtgctgtgttggaaagtaattggccatgagtggtggc	390
QY	2458	gagcaggagtgctgtcagggccccccagccccccagagagggcctgtgacatgggg	2517
Db	391	aagcaggagtgctgtcagggccccccagccccccagagagggcctgtgtgacatgggc	450
QY	2518	cctgtactactgctgtgttccagagcagctggagctccaa	2559
Db	451	ccctgtaccacagccctgtgttccagagcagctggagctccaa	492
RESULT	12		
BG005397			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
COMMENT			
FEATURES			
source			
BASE COUNT	110 a	181 c	192 g
ORIGIN			
Query Match	12.8%	Score 338	DB 10; Length 592
Best Local Similarity	84.6%	Pred. No. 2e-59	
Matches 391	Conservative	0	Mismatches 70; Indels 1; Gaps 1
QY	2099	tggatgaacgcagctgtgccc-gcgggtgccaggccccccagcctccctgaacctgcac	2157
Db	31	tggatgaacagagctgtgcccggggggccagagccccagctgtccctgaacctgcac	90
QY	2158	ggaccccatgcccccatactggagctggcagtgagacatccatccacccctctgt	2217
Db	91	agaccacccgtgtcccccatactggagctggcagtgagacatccatccacccctctgt	150
QY	2218	ggccccggcaccacagcagcagctgcagtcgcccggcagggaatttgggggggggtgcc	2277
Db	151	ggccccggcaccacagcagcagctgcagtcgcccggcagggaatttgggggggggtgcc	210
QY	2278	tcgggtgcccccagcgtgtgacatcccccggcccccaacatccacccagcttcgag	2337
Db	211	tcaggtccctccagagcgtgtgacatcccccggcccccaacatccacccagcttcgag	270
QY	2338	ctggcgcctctgtggccattgggaagtggcctctctcttgagccagctgcgcgtgc	2397
Db	271	ctggcgcctctgtggccattgggaagtggcctctctcttgagccagctgcgcgtgc	330

[illegible]



High quality sequence stop: 375.  
Location/Qualifiers  
1. .375  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0178"  
/dev\_stage="Adult"  
/note="Organ: placenta\_normal; Vector: puc18; Site\_1: SmaI  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
81 a 124 c 106 g 63 t 1 others

BASE COUNT 81 a 124 c 106 g 63 t 1 others  
ORIGIN

Query Match 12.1%; Score 317.8; DB 10; Length 375;  
Best Local Similarity 97.4%; Pred. No. 2.5e-55;  
Matches 333; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 312 ccagactctccagaaacccctccctttacagacacagctcctgggaagggtggccc 371  
Db 27 CCAGACTTCTCCAGAAACCTCCCTCC-TCTACAGGACACAGTCTCGGGGAAGGGTGGGCC 85

QY 372 acttcagaggtccggttcccaactaggagagagagagagagagagagagagagag 431  
Db 86 ACTTCAGGTCCCGTCCCACTAGGAGAGAGAGAGAGAGAGAGATTCAGAGCGGCGAG 145

QY 432 gagggtcccggtctcagaccctccatcaagccaggaagtctcggttatggagagtgccctt 491  
Db 146 GAGGTCCGTGCTCGAGACCCCTCAAGCCAGGAATGTCGGTATGGGAGAGTGCCCTT 205

QY 492 tgcattgcacatgcacccggaacccgagcagcctcctggagccacccagatctgagctgtc 551  
Db 206 GCGATGGCCACGGCACCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 265

QY 552 cctgatctctttagagggagagagcttccgtccctcactcaagagagagagagagag 611  
Db 266 CTGATCTCTTCTAGGGGAGAGACCTATTCTGTCCTCTACTCCAGAGGAGAGCCATT 325

QY 612 ctccgcaacgagcagcccccaactgagctccctcccccacaga 653  
Db 326 CTCGCAACGAGCGAGCGCCCAANTGAGTCCGCCCCACAGA 367

RESULT 15  
BF993027  
LOCUS BF993027 324 bp mRNA linear EST 23-JAN-2001  
DEFINITION IL5-GN0178-311000-204-e09 GN0178 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF993027  
VERSION BF993027.1 GI:12399350  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 324)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-GN0178-  
311000-204-e09&t3=2000-10-31&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 315.  
Location/Qualifiers  
1. .324  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0178"  
/dev\_stage="Adult"  
/note="Organ: placenta\_normal; Vector: puc18; Site\_1: SmaI  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
69 a 115 c 85 g 55 t

BASE COUNT 69 a 115 c 85 g 55 t  
ORIGIN

Query Match 11.9%; Score 314.4; DB 10; Length 324;  
Best Local Similarity 99.7%; Pred. No. 1.2e-54;  
Matches 315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 363 ggggtggccacttcgaggtcccgcttccacacccagagagagagagagagagagagag 422  
Db 1 GGTGGGCCACTTCGAGGTCCCGTTCGAGACCCCATCAAGCAGGAATGTCGGTTATGGGAG 60

QY 423 agcgccagagaggtcccggttcgagaccccatcaagccaggaatgttgggttatggag 482  
Db 61 AGCGGCCAGGAGGTCCCGGCTTCGAGACCCCATCAAGCAGGAATGTCGGTTATGGGAG 120

QY 483 agtgcctttgcatcctccactgcacggaaacccgagcaccctcgagccaccagatc 542  
Db 121 AGTCCCTTTTGCAATGCCACTGCACCGGAACCCAGCACCCTTCGGAGGCCACCCAGATC 180

QY 543 tgagctgcctctgatctcttctagagggggaagggtatttccctccctactccaagagc 602  
Db 181 TGAGCTGTCCTGATCTCTTCTAGAGGGGAAGAGCCTATTCCGTCCTCTACTCCAAGAGC 240

QY 603 agagccattctcgaacagcgagcccccacaaactgagctccctccacagaaactgtctgt 662  
Db 241 AGAGCCATTCTCGCAACCGCAGCCCCCAACTGAGCTCCCTCCACAGAACTGTCTGT 300

QY 663 ccacacccccatcccc 678  
Db 301 CCACACCCCATCCCC 316

Search completed: July 23, 2002, 16:20:47  
Job time: 11957 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 18:22:05 ; Search time 64.95 Seconds  
(without alignments)  
9961.488 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atgggaactggactggcag.....cctcagccttcacgcatag 2634

Scoring table: OLIGO.NUC  
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	0.7	410	3	US-09-114-146-1
2	19	0.7	410	5	PCT-US96-08623-1
3	19	0.7	2055	3	US-08-872-855-3
4	19	0.7	2800	3	US-08-872-855-1
5	18	0.7	1383	1	US-08-484-044-1
6	18	0.7	1787	2	US-08-372-652-6
7	18	0.7	1787	5	PCT-US95-16311-6
8	18	0.7	2511	2	US-08-422-699A-8
9	18	0.7	2511	2	US-08-422-706B-8
10	18	0.7	2607	2	US-08-907-166-1
11	18	0.7	2610	2	US-08-989-386-2
12	18	0.7	2726	2	US-08-422-699A-12
13	18	0.7	2726	2	US-08-422-706B-12
14	18	0.7	3182	1	US-08-484-044-11
15	18	0.7	3323	2	US-08-422-699A-10
16	18	0.7	3323	2	US-08-422-706B-10
17	18	0.7	4508	5	PCT-US93-06251-34
18	18	0.7	11613	1	US-08-484-044-10
19	17	0.6	300	3	US-09-157-177-117
20	17	0.6	427	2	US-08-623-906A-5
21	17	0.6	522	1	US-08-428-732-1
22	17	0.6	636	4	US-08-930-117-7
23	17	0.6	812	1	US-08-656-253-1
24	17	0.6	1581	1	US-08-313-553-8
25	17	0.6	1581	3	US-08-767-993-8
26	17	0.6	1683	1	US-08-164-614A-5
27	17	0.6	1683	2	US-08-456-489B-5

C 28	17	0.6	1683	5	PCT-US93-01720-5	Sequence 5, Appli
C 29	17	0.6	1758	1	US-08-078-222B-1	Sequence 1, Appli
C 30	17	0.6	1758	2	US-08-661-330A-1	Sequence 1, Appli
C 31	17	0.6	1758	3	US-09-038-217A-1	Sequence 1, Appli
C 32	17	0.6	1810	1	US-08-299-849B-20	Sequence 20, Appli
C 33	17	0.6	1810	2	US-08-142-388A-20	Sequence 20, Appli
C 34	17	0.6	1810	3	US-08-967-727-20	Sequence 20, Appli
C 35	17	0.6	1810	4	US-08-037-230D-20	Sequence 20, Appli
C 36	17	0.6	1892	3	US-08-620-077B-2	Sequence 4, Appli
C 37	17	0.6	1947	1	US-08-164-614A-4	Sequence 4, Appli
C 38	17	0.6	1947	2	US-08-456-489B-4	Sequence 4, Appli
C 39	17	0.6	1947	5	PCT-US93-01720-4	Sequence 4, Appli
C 40	17	0.6	1956	1	US-08-313-553-6	Sequence 6, Appli
C 41	17	0.6	1956	3	US-08-767-993-6	Sequence 6, Appli
C 42	17	0.6	1997	1	US-08-164-614A-6	Sequence 6, Appli
C 43	17	0.6	1997	2	US-08-456-489B-6	Sequence 6, Appli
C 44	17	0.6	1997	5	PCT-US93-01720-6	Sequence 6, Appli
C 45	17	0.6	2040	3	US-08-604-789B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-114-146-1  
; Sequence 1, Application US/09114146  
; Patent No. 6083747  
; GENERAL INFORMATION:  
; APPLICANT: WONG, Peter M.C.  
; APPLICANT: CHUNG, Siu-Wah  
; APPLICANT: HAN, Xiaodong  
; TITLE OF INVENTION: GLYCOPROTEIN GP105 ON BLS HEMATOPOIETIC  
; TITLE OF INVENTION: STEM CELLS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/114,146  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/471,188  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Best, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 46074/102/FEIN  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-114-146-1

Query Match 0.7%; Score 19; DB 3; Length 410;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Qy 1417 ggcctgtgagtcgtgtggg 1435  
Db 340 GGCCTGTGAGTCTGTGGG 358

RESULT 2  
PCT-US96-08623-1  
; Sequence 1, Application PC/TUS9608623  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: GLYCOPROTEIN GP105 ON BLS HEMATOPOIETIC  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/08623  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/471,188  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 46074/103/FEIN  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; PCT-US96-08623-1

Query Match 0.7%; Score 19; DB 5; Length 410;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1417 ggcctgtgagtcgtgtggg 1435  
Db 340 GGCCTGTGAGTCTGTGGG 358

RESULT 3  
US-08-872-855-3/c  
; Sequence 3, Application US/08872855  
; Patent No. 6121045  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean  
; TITLE OF INVENTION: NOVEL HUMAN DELTA3 COMPOSITIONS AND  
; TITLE OF INVENTION: THERAPEUTIC USES THEREFOR  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA

COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/872,855  
FILING DATE: 11-JUN-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MAA-003.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2055 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-872-855-3

Query Match 0.7%; Score 19; DB 3; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2404 gcccagagagccgcagc 2422  
Db 1417 GCCCAGAGAGCCGCGAGG 1399

RESULT 4  
US-08-872-855-1/c  
; Sequence 1, Application US/08872855  
; Patent No. 6121045  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean  
; TITLE OF INVENTION: NOVEL HUMAN DELTA3 COMPOSITIONS AND  
; TITLE OF INVENTION: THERAPEUTIC USES THEREFOR  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/872,855  
FILING DATE: 11-JUN-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MAA-003.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2800 base pairs



```

; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16311
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/372,652
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/246001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1787 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
PCT-US95-16311-6

Query Match 0.7%; Score 18; DB 5; Length 1787;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 690 tctaaagccctgaactgc 707
Db 37 TCTAAGCCCTGAAACTGC 20

RESULT 8
US-08-422-699A-8/C
; Sequence 8, Application US/08422699A
; Patent No. 5955265
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
; APPLICANT: Johnson, Keith J.
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02713
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,699A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,706
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/023,612

```

```

; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/839,255
; FILING DATE: 20-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01545
; FILING DATE: 19-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00253
; FILING DATE: 05-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB9202485.0
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5830A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1746
US-08-422-699A-8

Query Match 0.7%; Score 18; DB 2; Length 2511;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctcccccgc 263
Db 1920 GCCCCTCCCTCCCGGCC 1903

RESULT 9
US-08-422-706B-8/C
; Sequence 8, Application US/08422706B
; Patent No. 5977333
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
; APPLICANT: Johnson, Keith J.
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02713
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,706B
; FILING DATE: 14-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/284,543

```

```

Query Match      0.7%; Score 18; DB 2; Length 2607;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 834 ccctcagccacgaaggcc 851

```

```

RESULT 12
US-08-422-699A-12/c
; Sequence 12, Application US/08422699A
; Patent No. 5953265
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
;

```

APPLICANT: Johnson, Keith J.  
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,699A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/422,706  
FILING DATE:  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5830A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-422-699A-12

Query Match 0.7%; Score 18; DB 2; Length 2726;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctcccgccg 263  
|||||  
DB 2173 GCCCTCCCTCCCGGCC 2156

RESULT 13  
US-08-422-706B-12/c  
Sequence 12, Application US/08422706B  
Patent No. 5977333  
GENERAL INFORMATION:  
APPLICANT: Brook, J. David  
APPLICANT: Housman, David E.  
APPLICANT: Shaw, Duncan J.  
APPLICANT: Harley, Helen G.  
APPLICANT: Johnson, Keith J.

TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,706B  
FILING DATE: 14-APR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/284,543  
FILING DATE: 08-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5830A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-422-706B-12

Query Match 0.7%; Score 18; DB 2; Length 2726;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctcccgccg 263  
|||||  
DB 2173 GCCCTCCCTCCCGGCC 2156

RESULT 14  
US-08-484-044-11/c  
Sequence 11, Application US/08484044  
Patent No. 555282  
GENERAL INFORMATION:  
APPLICANT: Caskey, C. T.  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Friedman, David L.  
APPLICANT: Pizzuti, Antonio  
APPLICANT: Fenwick, Raymond G.  
TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy

NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski, L.L.P.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: Texas  
COUNTRY: U.S.A.  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,044  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/019,940  
FILING DATE: 19-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul, Thomas D.  
REGISTRATION NUMBER: 32,714  
REFERENCE/DOCKET NUMBER: D-5443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713/651-5325  
TELEFAX: 713/651-5246  
TELEX: 762829  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3182 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-484-044-11

Query Match 0.7% Score 18; DB 1; Length 3182;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 gccctccctcccccggcc 263  
|||||  
Db 2637 GCCCTCCCTCCCGGCC 2620

RESULT 15  
US-08-422-699A-10/c  
Sequence 10, Application US/08422699A  
Patent No. 5955265  
GENERAL INFORMATION:  
APPLICANT: Brook, J. David  
APPLICANT: Housman, David E.  
APPLICANT: Shaw, Duncan J.  
APPLICANT: Harley, Helen G.  
APPLICANT: Johnson, Keith J.  
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

Query Match 0.7% Score 18; DB 2; Length 3323;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 gccctccctcccccggcc 263  
|||||  
Db 2732 GCCCTCCCTCCCGGCC 2715

Search completed: July 23, 2002, 19:57:14  
Job time: 5709 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 17:26:55 ; Search time 2033.33 seconds  
(without alignments)  
17484.117 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1. atggagaactggactggcag.....cctcagcctttccagcatag 2634

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pln:\*
  - 16: em\_gss\_vrt:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	360	13.7	437	10	BM148104 TCAAP1Q83
2	328	12.5	370	10	EG008397 PMI-GN018
3	295	11.2	891	9	AL513944 AL513944
4	287	10.9	360	10	BF993048 IL5-GN017
5	286	10.9	307	10	BF989747 IL5-GN017
6	282	10.7	309	9	AI637480 tt07h07.x
7	285	10.1	324	10	BF993027 IL5-GN017
8	248	9.4	373	10	BF987707 QV0-GN014
9	246	9.3	507	10	BF986444 QV4-GN014
10	222	8.4	625	10	BF9838243 602490152
11	211	8.0	262	10	BF989357 IL5-GN017
12	205	7.8	711	10	BF989357 IL5-GN017
13	193	7.3	193	10	BF870887 IL0-ETO15
14	188	7.1	286	10	D78761 HUM510H03B
15	168	6.4	183	10	BF997839 MR2-GN012
16	167	6.3	214	10	BF997840 MR2-GN012
17	142	5.4	216	10	BF367876 RC4-GN003

18 110 4.2 449 10 BG900891  
19 107 4.1 865 10 BI094238  
20 103 3.9 375 10 EG008790  
21 94 3.6 327 10 BF993841  
22 87 3.3 389 10 BG008820  
23 86 3.3 266 10 BM150365  
24 86 3.3 273 10 BF997426  
25 83 3.2 419 10 D63308  
26 65 2.5 297 10 BF365794  
27 52 2.0 274 10 BM148104  
28 41 1.6 123 10 BF994270  
29 39 1.5 125 10 BG001393  
30 39 1.5 495 9 AW230557  
31 39 1.5 513 9 AW763191  
32 32 1.4 127 10 BI012353  
33 38 1.4 390 10 BF368874  
34 34 1.3 597 9 AI683106  
35 32 1.2 592 10 BM391602  
36 32 1.2 866 10 BI330332  
37 30 1.1 517 10 BE808908  
38 29 1.1 480 10 BF651524  
39 29 1.1 571 10 BE757541  
40 28 1.1 188 10 BG242515  
41 26 1.0 360 9 AW482344  
42 26 1.0 379 9 BB857357  
43 26 1.0 446 9 BB862174  
44 26 1.0 483 9 BB856356  
45 26 1.0 498 9 BB855583

RESULT 1  
BM148104  
LOCUS  
DEFINITION  
TCAAP1Q8352 Pediatric acute myelogenous leukemia cell (FAB M1)  
Baylor-HGSC project-TCAA Homo sapiens CDNA clone TCAAP8352, mRNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BM148104 437 bp mRNA linear EST 30-NOV-2001  
TCAAP1Q8352 Pediatric acute myelogenous leukemia cell (FAB M1)  
Baylor-HGSC project-TCAA Homo sapiens CDNA clone TCAAP8352, mRNA  
sequence.  
BM148104 1 GI:17168519  
EST.  
BM148104.1  
human.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 437)  
Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman, F.R. Jr.,  
Gunaratne, P.H., Muzny, D., Bouck, J., Gibbs, R.A. and Margolin, J.F.  
Pediatric Leukemia CDNA Sequencing Project (2001)  
Unpublished (2001)  
Contact: Dr. Judith F. Margolin  
Texas Children's Cancer Center and Human Genome Sequencing Center  
at Baylor College of Medicine  
1102 Bates, MC3-3320 Houston, TX 77030, USA  
Tel: 832-824-4536  
Fax: 832-825-4038  
Email: clones@txccc.org  
Seq primer: M13 primer.  
Location/Qualifiers  
1. 437  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TCAAP8352"  
M1) Baylor-HGSC project-TCAA"  
/sex="male"  
/tissue\_type="leukopheresis"  
/cell\_type="myeloid cell"  
/dev\_stage="pediatric 6 years"  
/lab\_host="DH10B"  
/note="vector: lambda pSB; Site\_1: BamHI; Site\_2: EcoRI;  
First strand cDNA was primed with an anchored

ALIGNMENTS





ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 360)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F., F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel.: +55-11-2704922

```

Email: asimpson@ludwig.org.br
Fax: +55-11-2707001
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&st2=IL5-GN0178-
311000-204-f10&t3=2000-10-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 305.
Location/Qualifiers
    1..360
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="GN0178"
    /dev_stage="Adult"
    /note="Organ: placenta_normal; Vector: puc18; Site.1: SmaI
    ; Site.2: SmaI; A mini-library was made by cloning
    products derived from ORESTES PCR (U.S. Letters Patent
    application No. 196,716 - Ludwig Institute for Cancer
    Research) profiles into the pUC 18 vector. Reverse
    transcription of tissue mRNA and cDNA amplification were
    performed under low stringency conditions."
    59 a      80 c      142 g      78 t      1 others
FEATURES
    source
        Query Match          10.9%; Score 287; DB 10; Length 360;
        Best Local Similarity 99.7%; Pred. No. 3.9e-126;
        Matches 337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
        492  tgcattgcacatgcacgcgaaccgcaggacccctgcggagccaccacagatctgagctgtc 551
        538  tgcattgcacatgcacgcgaaccgcaggacccctgcggagccaccacagatctgagctgtc 579
        552  cctgatctctctagagggggaagagctattccgtccctactccaagacagagccatt 611
        578  cctgatctctctagagggggaagagctattccgtccctactccaagacagagccatt 619
        612  ctccgcaaacgcagcccccaactgacctccctccacagaaactgtctgtccacacccc 671
        618  ctccgcaaacgcagcccccaactgacctccctccacagaaactgtctgtccacacccc 659
        672  atcccccaagcagaacctctaagccctgaaactgtctcagacagaggtggccccccagaac 731

```

732 caggcctccccctacggcatcacccacagagccaggctctaggcacagagccccctc 791  
9 98 CAGGCTGCCCCCTACGCGCATCACCCAGAGCCCCAGGCTCTGGCAGAGCCCCCTC 39  
792 accacgcactccttaggaaggtggctcttcggtg 829  
9 38 ACCCAGCGACCTCTTAGGAGAGGTGGCTCTTCGGTG 1

QY	503	agagccatttcgaaaacgagcccccaactgagctccctccc	648
DB	241	AGAGCCATTCTCCGCAACGGCAGCCCCCAAGCTAGCTCCCTCCC	286
RESULT	6		
LOCUS	AI637480	309 bp	mRNA linear EST 14-DEC-1999
DEFINITION	t07n07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2240125 3' similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ; mRNA sequence.		
ACCESSION	AI637480		
VERSION	AI637480.1	GI:4689714	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/hbrp/image/image.html Insert length: 409 Std Error: 0.00 Seq primer: -40UP from Gibco. Location/Qualifiers		
FEATURES			

```

1. .309
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:240125"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10s"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743)."

```

```

BASE COUNT      56 a      92 c      91 g      70 t
ORIGIN
Subtraction by Bento Soares and M. Fatima Bonaldo.

Query Match      10.7%; Score 282; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 9,6e-124;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2353 cattggaagttagcttcctctggcgcagtgctccgtgcggtgcggtgcgcgggcccagaga 2412
      |||||
DB 309 CATTTGGGAAGTTGGCTTCCTTTGGAGCCAGTGTCTCGTTCGGTGCCTGGCGCGGGCCAGAGA 250
      |||||

QY 2413 agcggcaggttgcgtgtgttgggaacaacggtgatgaagtgcgagcagcaggtgtgcg 2472
      |||||
DB 249 AGCGGGCAGGTTCTGCTGTGTGGGAACAACCGTGTATGAAGTGCAGCGCAGAGGTGTGCG 190
      |||||

QY 2473 tcaggccccccacagccccccagcagagagccctgtgacatggggccctgtactactgcc 2532
      |||||
DB 189 TCAGGCCCCCCACAGCCCCCCCCAGCAGAGGCCCTGTGACATGGGGCCCTGTACTACTGCC 130
      |||||

```

```

QY 2533 tgggtccacagcactggagctccagggtagcccgagcccccacccacccatattctgcatc 2592
|||||
Db 129 TGGTTCCACAGCGACTGGAGCTCCAGGTGAGCCCGGAACCCCGACCCATATCTCTGCATC 70
|||||

QY 2593 ctgggttaaccatccacagacactcagcctttccagcatag 2634
Db 69 CTGGGTAAACCATCCCGCAGGACACTCAGCCTTCCAGCATAG 28
|||||

RESULT 7
BF993027
LOCUS IL5-GN0178-311000-204-e09 GN0178 Homo sapiens cDNA, mRNA sequence. EST 23-JAN-2001
DEFINITION BF993027
ACCESSION BF993027
VERSION BF993027.1 GI:12399350
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 324)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t3=IL5-GN0178-
311000-204-e09&t3=2000-10-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 315.
Location/Qualifiers
1..324
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0178"
/dev_stage="Adult"
/note="Organ: placenta_normal; Vector: puc18; Site_1: Smai
; Site_2: Smai; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 69 a 115 c 85 g 55 t
ORIGIN

Query Match 10.1%; Score 265; DB 10; Length 324;
Best Local Similarity 99.7%; Pred. No. 1.3e-115;
Matches 315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 363 gggtggccacttcaggtcccgctccacccatagggagagagagagagagattcg 422
|||||
Db 1 GGGTGGCCACATCTCGAGTCCCGCTCCACCTAGGAGAGAGAGAGAGAGAGATCG 60
|||||

QY 423 agcgccagagaggtcccgctcccgagacccatcagccaggaatttcgttatggag 482
|||||
Db 61 AGCGCCAGAGAGTCCCGCTCCGAGACCCCATCAAGCCAGGAGATGTTCCGTTATGGGAG 120
|||||

```

```

QY 483 atgcccattgcatggccacitgacccggaacccgagccacccctggagcccccacagatc 542
|||||
Db 121 AGTGCCCTTTGATGTGACATGACACCGAACCAGGACACCTCGAGAGCCACCCAGATC 180
|||||

QY 543 tgagctgccctgatctcttctagagggagaggtctattccgtccctactcccaagagc 602
|||||
Db 181 TGAGCTGTCCCTGATCTCTCTAGAGGGGAGAGAGCTATTTCGTCCTCCCTACTCCAGAGC 240
|||||

QY 603 agagccattctccgaaacgagcagcccccacaaactgactccctccacagaaactctgt 662
|||||
Db 241 AGAGCCATTCTCCGAAACGGCAGCCGCCCAACTGAGTCTCCCTCCACAGAACTCTCTGT 300
|||||

QY 663 ccacaccccccatcccc 678
Db 301 CCACACCCCATCCCCC 316
|||||

RESULT 8
BF987707
LOCUS QV0-GN0140-121000-430-b09 GN0140 Homo sapiens cDNA, mRNA sequence. EST 23-JAN-2001
DEFINITION BF987707
ACCESSION BF987707
VERSION BF987707.1 GI:12394029
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 373)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-GN0140-
121000-430-b09&t3=2000-10-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 75
High quality sequence stop: 373.
Location/Qualifiers
1..373
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0140"
/dev_stage="Adult"
/note="Organ: placenta_normal; Vector: puc18; Site_1: Smai
; Site_2: Smai; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 81 a 125 c 98 g 69 t
ORIGIN

Query Match 9.4%; Score 248; DB 10; Length 373;

```



```

BASE COUNT      188 a      161 c      174 g      102 t
ORIGIN

Query Match      8.4%; Score 222; DB 10; Length 625;
Best Local Similarity 100.0%; Pred. No. 5.4e-95;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1464 ggggaacctcactgacgagggcccccctggcgtatcagaagatcttggattccagc 1523
      |||
Db 51 GGGGAACCTCAGTACCGAGGGGGCCCCCTGGGCTATCAGAAATCTTGTGGATTCCAGC 110
      |||

QY 1524 gggagccttggcgtccagattgccagctccggcctagctccaaactacctggcacttgg 1583
      |||
Db 111 GGGAGCCTTGGCGTCCAGATTGCCAGCTCCGGCTAGTCCCAACTACCTGGCACTTCG 170
      |||

QY 1584 tggccctggggcgggtccatcaatggaactgggctgtggtatcccccctggctccta 1643
      |||
Db 171 TGGCCCTGGGGGGCGGTCCTATCATCAATGGGAACGTGGGCTGTGGATCCCTGGTCCCTA 230
      |||

QY 1644 caggccgcggggagccgtcttctgatataaccgtctcccccag 1685
      |||
Db 231 CAGGCGCGGGGACCGCTCTTCGATATACCGTCTCCCCAG 272
      |||

RESULT 11
LOCUS      BF989357/262 bp mRNA linear EST 23-JAN-2001
DEFINITION IL5-GN0179-201000-185-e11 GN0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BF989357
VERSION      BF989357.1 GI:12395682
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 262)
AUTHORS      Dias Neto, E., Garcia Correa R., Verjovski-Almeida, S., Briones, M.R.,
               Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
               Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
               Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
               M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
               Simpson, A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5st2-IL5-GN0179-
               201000-185-e11st3-2000-10-20st4=1)
               Seq primer: puc 18 forward
               High quality sequence stop: 262.
               Location/Qualifiers
                 1..262
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="GN0179"
                   /dev_stage="Adult"
                   /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                   ; Site_2: SmaI; A mini-library was made by cloning
                   products derived from ORESTES PCR (U.S. Letters Patent
                   application No. 196,716 - Ludwig Institute for Cancer
                   Research) profiles into the pUC 18 vector. Reverse
                   transcription of tissue mRNA and cDNA amplification were
BASE COUNT      56 a      86 c      79 g      41 t
ORIGIN

Query Match      8.0%; Score 211; DB 10; Length 262;
Best Local Similarity 99.6%; Pred. No. 9.1e-90;
Matches 261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1432 gggggtgatgttacctgctgccttgttcggggaaacctcactgacgagggggcccc 1491
      |||
Db 262 GGGGGTGATGATCTTACCTGTGCCTTGTTCGGGGAACTCCTACTGACCGAGGGGGCCCC 203
      |||

QY 1492 ctgggctatcagaagatcttggattccagcggagccttgcggctccagattgccag 1551
      |||
Db 202 CTGGGGTATCAGAAGATCTTGTGGATTCCAGCGGGAGGCTTCGGCTCCAGATTCCCGAG 143
      |||

QY 1552 ctccggctcgtcacaactacctgctgaccttgccttggggcggcggtccatcaat 1611
      |||
Db 142 CTCGGCCCTAGCTCTCCACTTACCTGGCATTCTGTGCCCTGGGGCGGGTCCATCATCAT 83
      |||

QY 1612 ggggaactgggctgtggatccccctgggtctctacagggcgcgggacccgttttcgat 1671
      |||
Db 82 GGGAACTGGGCTGTGGATCCCGTGGGTCTTACAGGCGCGGGACCGTCTTTCGATAT 23
      |||

QY 1672 aacctctcccccaggaggagg 1693
      |||
Db 22 AACCGTCTCCCGAGGAGGAGG 1

RESULT 12
LOCUS      BI836313/711 bp mRNA linear EST 04-OCT-2001
DEFINITION 603083995F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5223358 5',
               mRNA sequence.
ACCESSION      BI836313
VERSION      BI836313.1 GI:15947863
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 711)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
               Email: cgabbs@mail.nih.gov
               Tissue Procurement: Life Technologies, Inc.
               CDNA Library Preparation: Life Technologies, Inc.
               CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM11561 row: m column: 23
               High quality sequence stop: 711.
               Location/Qualifiers
                 1..711
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:5223358"
                   /clone_lib="NIH_MGC_120"
                   /lab_host="DH10B"
                   /note="Organ: pooled pancreas and spleen; Vector:
                   pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                   source anonymous pool of spleen and pancreas from 28 yo
                   male. Library is oligo-dT primed and directionally cloned
                   (EcoRV site is destroyed upon cloning). Average insert
                   size 1.5 kb, insert size range 1-2.5 kb. Library is
                   normalized and enriched for full-length clones and was
                   constructed by C. Gruber (Invitrogen). Research Genetics
                   tracking code 025. Note: this is a NIH_MGC Library."
FEATURES
source

```

```

BASE COUNT      151 a      233 c      222 g      104 t      1 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 205; DB 10; Length 711;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2178 ctgggaagctggagtgagacatctctgacgcctctctgtggccccggcaccagacacg 2237
Db 210 CTGGAGGCTGGGAGTGGACATCTCTGACGCCCTCTGTGGCCCCGGCACCCAGACCG 151

QY 2238 ccagctcagtgccggcaggaaattgggggggtgggtctctctgtgccccggagcgctg 2297
Db 150 CCAGCTCAGTGGCGGAGGAATTTGGGGGGTGGCTCTCGTGCCCGCGAGCGCTG 91

QY 2298 tggacatctccccggcccaacatcacccagttgttgcagctgcgctctgtggccattg 2357
Db 90 TGGACATCTCCCCGGGCCCAACATCACCCAGTCTTGCCAGCTGCGCTCTGTGGCCATTG 31

QY 2358 ggaagtggctctctctggagccag 2382
Db 30 GGAAGTTGGCTCTCTCTGGAGCCAG 5

RESULT 13
BF870887
LOCUS BF870887 193 bp mRNA linear EST 17-JAN-2001
DEFINITION IL0-ET0152-301000-501-c03 ET0152 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF870887
VERSION BF870887.1 GI:12261017
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL0&t2=IL0-ET0152-
301000-501-c03&t3=2000-10-30&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 193.
FEATURES
source
1..193
Location/Qualifiers
/db_xref="taxon:9606"
/clone_lib="ET0152"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products
derived from ORFESTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under

```

---

```

BASE COUNT      25 a      50 c      68 g      50 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 193; DB 10; Length 193;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1342 cctgacatctgtgtggctgagacgtgtctgagccccgctgtgagatccttgctct 1401
Db 1 CCTGACATCTGTGTGGCTGGACGCTGTCTGAGCCCGGCTGTGATGGGATCCTTGGCTCT 60

QY 1402 gscaggctccctgagctgtgagctgtggtggtgagcttacctgtcgctctgtt 1461
Db 61 GCAGGCGTCCCTGATGCGCTGTGGAGTCTGTGGGGTGAATTCCTACCTGTGCGCTTGT 120

QY 1462 tggggaaacctcactgacgagggggccccctgtggtctatcagaagatttgggtacca 1521
Db 121 TCGGGAACTCACTACGACGAGGGGGCCCCCTGGGCTATCAGAGATCTTGGATTCCA 180

QY 1522 gcggaagccttgc 1534
Db 181 GCGGGAGCCTTGC 193

RESULT 14
D78761
LOCUS D78761 286 bp mRNA linear EST 09-FEB-1996
DEFINITION HUM510H03B Human placenta polyA+ (TFujiwara) Homo sapiens cDNA
clone GEN-510H03 5', mRNA sequence.
ACCESSION D78761
VERSION D78761.1 GI:1180634
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 286)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi
A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H.,
Shin,S. and Nakamura,Y.
TITLE Fujiwara et al. (1995)
JOURNAL Unpublished (1995)
COMMENT Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd
463-10 kagasuno Kawauchi-cho, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.
FEATURES
source
1..286
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-510H03"
/clone_lib="Human placenta polyA+ (TFujiwara)"
/tissue_type="placenta"
BASE COUNT      49 a      91 c      93 g      53 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 188; DB 10; Length 286;
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2287 cgggaagctgtggacatctccccggggcccaacatcacccagtttgcagctgcctc 2346
Db 1 CCGGAGCGTGTGGACATCTCCCGCGGCCAACATCACCCAGCTGTGCCAGCTGCGCTTC 60

QY 2347 tgtggccattggaaattggctctctcttggagccagtgctccgtcggtgctggccggggc 2406
Db 61 TGTGGCCATTGGGAATTTGGCTCTCCTTTGGAGCCAGTGTCTCCGTGCGGCGCGGGGC 120

```



Db 76 CATTGGGAAGTTGGCTCTCCTTTGGAGCCAGTCTCCCTGGGTCGGCGGCCGAGAGA 135  
Qy 2413 agccggcaggttcgctgtgttgggaacacacggtgatgaagtgcgag 2460  
Db 136 AGCCGGCAGGTTGCTGTGTGGGAACAACGGTGATGAAGTGAGCGAG 183

Search completed: July 23, 2002, 19:00:50  
Job time: 5635 sec

Qy 2407 cagagaagccgaggttcgctgtgttgggaacacacggtgatgaagtgcgagcagag 2466  
Db 121 CAGAGAAGCCGCGAGTTCGCTGTGTGGGAACAACGGTGATGAAGTGAGCGAGCAGAG 180  
Qy 2467 tgtgcgtc 2474  
Db 181 TGTGGGTC 188

RESULT 15  
BF997839  
LOCUS BF997839 183 bp mRNA linear EST 23-JAN-2001  
DEFINITION MR2-GN0127-091100-005-402 GN0127 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF997839  
VERSION BF997839.1 GI:12404162  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 183)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

FEATURES  
source  
1. .183  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0127"  
/dev\_stage="Adult"  
/note="Organ: placenta normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
BASE COUNT 28 a 56 c 63 g 35 t  
ORIGIN

Query Match 6.4%; Score 168; DB 10; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.4e-69;  
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2293 cgcgtggagatctcccccgcacacatccacagcttcgagctgcctctgtgac 2352  
Db 16 CGCTGTGGACATCTCCCGCCGCCAACATCACCAGCTTGGCCAGCTGCGCCCTGTTGGC 75

Qy 2353 cattgggaagtgtgctctctcttgagccagtgctcccgcggtgcgagccagaga 2412



```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      July 23, 2002, 13:08:30 ; Search time 65.15 Seconds
              (without alignments)
              9930.908 Million cell updates/sec

Title:      US-10-041-770-1
Perfect score:      2634
Sequence:      1 atggagaactggactggag.....cctcagcctttccagcatag 2634

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:      383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters:      767066

```

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database : .
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	48.8	1.9	4403765	4	US-09-103-840A-2		Sequence 2, Appli
2	48.8	1.9	4411529	4	US-09-103-840A-1		Sequence 1, Appli
3	47.2	1.8	152331	3	US-09-128-155-16		Sequence 15, Appl
c 4	43.2	1.6	400	4	US-09-056-556-179		Sequence 179, App
5	43.2	1.6	7218	1	US-08-232-463-14		Sequence 14, Appl
c 6	42.2	1.6	1443	1	US-08-076-089-1		Sequence 1, Appli
c 7	42.2	1.6	1443	2	US-08-707-200-1		Sequence 1, Appli
c 8	42.2	1.6	1443	4	US-08-956-565-1		Sequence 1, Appli
c 9	42.2	1.6	1443	5	UCT-US93-05643-1		Sequence 1, Appli
c 10	41.6	1.6	4403765	4	US-09-103-840A-2		Sequence 2, Appli
c 11	40.2	1.5	320	4	US-09-165-364-11		Sequence 11, Appl
c 12	40.2	1.5	834	4	US-08-998-416-437		Sequence 437, App
13	39.4	1.5	595	4	US-08-483-533-4		Sequence 4, Appli
14	39.4	1.5	595	4	US-09-283-471A-4		Sequence 4, Appli
15	39.4	1.5	1327	4	US-08-483-533-36		Sequence 36, Appl
16	39.4	1.5	1327	4	US-09-283-471A-36		Sequence 36, Appl
17	39.4	1.5	1335	5	UCT-US91-06532-1		Sequence 1, Appli
18	38.4	1.5	1035	1	US-07-601-094-30		Sequence 30, Appl
19	38.4	1.5	1035	1	US-08-012-735-30		Sequence 30, Appl
20	38.4	1.5	1910	4	US-05-593-711A-3		Sequence 3, Appli
21	38.4	1.5	1914	1	US-07-601-094-1		Sequence 1, Appli
22	38.4	1.5	1914	1	US-08-012-735-1		Sequence 1, Appli
23	37.8	1.4	801	2	US-08-770-379-16		Sequence 16, Appl
24	37.8	1.4	801	4	US-08-757-669A-16		Sequence 16, Appl
25	37.8	1.4	801	4	US-05-230-371A-16		Sequence 16, Appl
c 26	37.8	1.4	914	2	US-08-935-450-10		Sequence 10, Appl
c 27	37.8	1.4	1280	4	US-09-060-756-4		Sequence 4, Appli



[illegible][illegible]

```

; FILING DATE: August 20, 1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/076,089
; FILING DATE: June 11, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/901,710
; FILING DATE: June 19, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: JDP-015CPDV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1443
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-707-200-1

Query Match          1.6%; Score 42.2; D8 2; Length 1443;
Best Local Similarity 44.4%; Pred.No.0.39;
Matches 170; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY   17 gcaggccctggcgtatctgctgctgtcttgccttcctccctcaagctctgcttggaacagg 76
Db   412 GCAGCAGCACACCTTGTAAACGCTTCGCTCAGTCGCTGCCCTCGAGTCAGCGAGTGCT 353
QY   77 agtgttttcggcacactctctcagacacctcagagagagcgccagggccccgaagtq 136
Db   352 CGAGTCTCTCGGGCCAGTCCAGCTGGGACCCTCTGGGTCCCCTCCGCCGCCGGCGCGCG 293
QY   137 tctggggaccctgggttcagtgggcctctctgtcccagccccctgcggggtgagggtgcagc 196
Db   292 TCAGGGGCACCCGGGGGTGAGCGCCGCTCGAGCTCGAGTCGCGCTGCCACCGGCATGCTCGGC 233
QY   197 gcaggagcgcggacatgtcagctccctacagtgcagctccacccgagcttgccccctccctc 256
Db   232 GGTGCAGCGCGGGCGGGGCCCGACGAGGTGTGTCGCCCGACGCGCTCGCTCTCTGCG 173
QY   257 cccggccccaaagacatccagaagccctctcccccggggcgaggggtccacagaccocaga 316
Db   172 CCCCCACCGCGCTCCCGCCGCTCCGCTCGCGCGCGCTCAGGGTCAATCGCGTCCGGGA 113
QY   317 ctctccagaaacctccctcttgtaagggacacagctctcgggggaaggggtggcccaattc 376
Db   112 CGTCCACTCGCACACAGACAGCGCTCAGGACTAGGATCCGGATTAGGAGGCCACGACCTC 53
QY   377 gaggttcggcgttccacactaggg 399
Db   52 GTAGTCGGGACTCGGACTTTAGG 30

RESULT      8
US-08-996-565-1/c
; Sequence 1, Application US/08996565
; Patent No. 6326141
; GENERAL INFORMATION:
; APPLICANT: C. Ronald Kahn, M.D.
; APPLICANT: Christine Reynet, Ph.D.
; TITLE OF INVENTION: DIABETOGENE RAD: A TYPE II DIABETES
; TITLE OF INVENTION: SPECIFIC GENE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
```

```

; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 00303/015002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1443
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-076-089-1

```

Query Match 1.6%; Score 42.2; DB 1; Length 1443;  
 Best Local Similarity 44.4%; Pred. NO. 0.39;  
 Matches 170; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

QY 17 gcaggccctggctgtatctgctgtgtcttctccctccctcagctctcttggtgatcagg 76
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 GCAGCAGCACACCTTGTAAAGCTGCTCGCTGAGTCGCTGCCCTCGAGCTGAGCGAGTCCT 353
QY 77 agtgttttcgggacatctcttcagacacatcacagagagggcccgcccaagtg 136
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 CGAGTCTCTCGGGCAGTCCAGCTGGGACCCCTGGGTGCCCGCTCCGGCGCGCGCGCG 293
QY 137 tctggggacctgggttcagtgggccctcttctccacagccctgcggggtgggggtgcagc 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 TCAGGCGCACCGGGGTGTCAGGCGCCCTGCAAGTCGGCTGCTCCACCGGCATGCTGCGGC 233
QY 197 gcaggagcggacatgtcagctccctacatgcagctccacccagctctgccctccctc 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 GGTGAGAGGGGGGGGGGGCCCGAGGGTGTGTGTCGCCGAGGGCGTCGCTCCTGCGC 173
QY 257 cccggcccccagaagatctcagaagccctctccccccggggccagggtccccagaccccgaga 316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 CCCCACCGCGGCTCCGCGCCGCTCCGCTGCCGCGCGCTTCAGGGTCAATCGCGTCCGGGA 113
QY 317 ctctccagaagaccctcccttgtagaggacacagctctcggggaaaggggtgcccacttc 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 CGTCCACTCGCACACACAGCTCAGGATAGGATCCGGATTAGGAGGCCACGCACCTC 53
QY 377 gagggtccgcgttccacactaggg 399
    ||||| ||||| ||||| |||||
Db 52 GTAGTCCGAGACTCGGACTTTAGG 30

```

RESULT 7  
 US-08-707-200-1/G  
 ; Sequence 1, Application US/08707200  
 ; Patent No. 5851430  
 ; GENERAL INFORMATION:  
 ; APPLICANT: C. Ronald Kahn, M.D.  
 ; APPLICANT: Christine Reynet, Ph.D.  
 ; TITLE OF INVENTION: DIABETOGENE RAD: A TYPE II DIABETES  
 ; TITLE OF INVENTION: SPECIFIC GENE  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LAHIVE & COCKFIELD  
 ; STREET: 60 State Street  
 ; CITY: Boston  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02109-1875  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; COMPUTER: IBM PS/2 Model 502 or 55SX  
 ; OPERATING SYSTEM: MS-DOS (version 5.0)  
 ; SOFTWARE: WordPerfect (version 5.1)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08707,200



Query Match	1.5%	Score 40.2;	DB 4;	Length 320;
Best Local Similarity	49.8%;	Pred. No. 0.7;	Mismatches 0;	Indels 0; Gaps 0;
Matches 102;	Conservative			
Qy	1777	gttttttatcagatgcatcttcttcacctctccaactcttgagaaccccccacagay	1836	
DB	319	gncrtcttttcaaaattctgcctcccccccgcccccccgcccccccgcccccccgcccc	260	

Query Match	1.5%	Score 40.2	DB 4	Length 834
Best Local Similarity	56.4%	Pred. No. 0.97		
Matches	75	Conservative 0	Mismatches 58	Indels 0
Gaps				
QY	144	accttgggtccagtgggcctcttctccacgcctcggggtgggggtcagcgacgag	203	
DB	176	ATCATTCGCCAGATGTCFCGCTCCAGATGGGATGGCGGTGCGGCGCCGACGAGGC	235	

Db 242 CCCGCGAGCCCGCGGGTGGCGTTCTCCGCCACGTCCGGGTGCGCACCTGGTGGTC 301  
Qy 342 caggacacagtcctcggggaagggtggccacacttcaggtccgc 386  
Db 302 TGGGCTCGGCGCGCCCTGGCGCGCGGCTGCTGGGCGCGC 346

RESULT 14  
US-09-283-471A-4  
; Sequence 4, Application US/09283471A  
; Patent No. 6340673  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Chou, Joany  
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60608-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/09/283,471A  
; FILING DATE: 04-APR-1999  
; CLASSIFICATION: 514  
; PRIOR APPLICATION NUMBER: 07/861,233  
; FILING DATE: 31-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/419,853  
; FILING DATE: 11-APR-1995  
; PRIOR APPLICATION NUMBER: 08/483,533  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, James P.  
; REGISTRATION NUMBER: 28,491  
; REFERENCE/DOCKET NUMBER: 27373/32742A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 595 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-283-471A-4

Query Match 1.5%; Score 39.4; DB 4; Length 595;  
Best Local Similarity 48.4%; Pred. No. 1.4; Mismatches 0; Gaps 0;  
Matches 109; Conservative 0; Indels 116; Indels 0; Gaps 0;  
Qy 162 ctcttctccagccctgcggggtgagcgagagccgagacatgtcagctccc 221  
Db 122 CGCTCGAGCGCGGGGAGGGGCGCGGAGCCCGCGAGCCCGCGAGCCCGC 181  
Qy 222 tacagtgcagctccaccagctgtgcccctccctcccgcccaagacatccagaac 281  
Db 182 GCGACCCCGCGAGCCCGCGAGCCCGCGAGCCCGCGAGCCCGCGAGCCCGC 241  
Qy 282 cctctcccccggggccagggtccagagcccccagacttctccagaaacccctctgta 341

Qy 204 ccggacatgtcagctccctacagtcagctccaccgagtgctgcccctccctcccgagcc 263  
Db 236 ACCGACTTTCACGCCCGCGGTCCCGCCACACCGGAGCGCGCGCCCTGCGCGGAAC 295  
Qy 264 cccaagacatcca 276  
Db 296 CCGTGGCTGGCCA 308

RESULT 13  
US-08-483-533-4  
; Sequence 4, Application US/08483533  
; Patent No. 6172047  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Chou, Joany  
; TITLE OF INVENTION: Method for Treating Tumorigenic  
; TITLE OF INVENTION: Diseases  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/483,533  
; FILING DATE: 07-MAR-95  
; CLASSIFICATION: 514  
; PRIOR APPLICATION NUMBER: 08/419,853  
; FILING DATE: 11-APR-95  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/861,233  
; FILING DATE: 31-MAR-92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, James P.  
; REGISTRATION NUMBER: 28,491  
; REFERENCE/DOCKET NUMBER: 28097/32742  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 595 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-483-533-4

Query Match 1.5%; Score 39.4; DB 4; Length 595;  
Best Local Similarity 48.4%; Pred. No. 1.4; Mismatches 0; Gaps 0;  
Matches 109; Conservative 0; Indels 116; Indels 0; Gaps 0;  
Qy 162 ctcttctccagccctgcggggtgagcgagagccgagacatgtcagctccc 221  
Db 122 CGCTCGAGCGCGGGGAGGGGCGCGGAGCCCGCGAGCCCGCGAGCCCGC 181  
Qy 222 tacagtgcagctccaccagctgtgcccctccctcccgcccaagacatccagaac 281  
Db 182 GCGACCCCGCGAGCCCGCGAGCCCGCGAGCCCGCGAGCCCGCGAGCCCGC 241  
Qy 282 cctctcccccggggccagggtccagagcccccagacttctccagaaacccctctgta 341

DB 242 CCCGGACCCCGCGGGTGGCTTCTCGCCACACGTCGGGTGGCGCACCTGGTGGTC 301  
QY 342 caggacacagtcctcgagggaagggtggcccaacttcagaggtccgcg 386  
DB 302 TGGGCTCGGCGCGCGCTGGCGCGCGCGGCTCGTGGGCGCGC 346

## RESULT 15

US-08-483-533-36  
; Sequence 36, Application US/08483533  
; Patent No. 6172047  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Chou, Joany  
; TITLE OF INVENTION: Method for Treating Tumorigenic  
; TITLE OF INVENTION: Diseases  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,533  
; FILING DATE: 07-MAR-95  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/419,853  
; FILING DATE: 11-APR-95  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/861,233  
; FILING DATE: 31-MAR-92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, James P.  
; REGISTRATION NUMBER: 28,491  
; REFERENCE/DOCKET NUMBER: 28097/32742  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1327 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-483-533-36

Query Match 1.5%; Score 39.4; DB 4; Length 1327;  
Best Local Similarity 48.4%; Pred. No. 1.8;  
Matches 109; Conservative 0; Mismatches 116; Indels 0; Gaps 0;  
QY 162 ctctgtctccagccctcggtgggtgagcagcaggagccgacatgtcagctccc 221  
DB 563 CGCTGGACCGCGGGCGGAGGGGCGCGGAGCCCGCGGACCCCGCGGACCC 622  
QY 222 tacagtgagtcacccagtgctgcccctccctcccgcccaagacatccagaagc 281  
DB 623 GCGACCCCGCGGACCCCGGAGCCCGCGGAGCCCGCGGACCCCGCGGACCC 682  
QY 282 cctctcccccggggggggtccagaccccgagactctccagaacccctcttga 341  
DB 683 CCCGCGACCCCGCGGGGTGCGCTTCTCGCCCGGAGTCCGGGTGGCGCACCTGGTGGTC 742

QY 342 caggacacagtcctcgagggaagggtggcccaacttcagaggtccgcg 386  
DB 743 TGGGCTCGGCGCGCGCTGGCGCGCGGCTCGTGGGCGCGC 787

Search completed: July 23, 2002, 20:17:45  
Job time: 25755 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 20:17:51 ; Search time 24.4 Seconds  
(without alignments)  
877.920 Million cell updates/sec

Title: US-10-041-770-2  
Perfect score: 4895  
Sequence: 1 MENWGRPWLYLLLSLPQ.....PPAISILGNHAQDTSAPFA 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pap: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pap: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pap: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pap: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pap: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	284	5.8	788	2	US-08-918-914-4
2	232.5	4.7	1064	1	US-08-642-255-62
3	224.5	4.6	960	4	US-09-219-849-5
4	224.5	4.6	1185	4	US-09-041-886-23
5	220	4.5	882	4	US-09-413-814-78
6	216.5	4.4	1341	3	US-08-963-825-18
7	216.5	4.4	1341	4	US-09-570-573-18
8	216.5	4.4	1341	4	US-09-548-608-18
9	213	4.4	1055	1	US-08-642-255-72
10	210.5	4.3	1057	3	US-08-931-820-1
11	210.5	4.3	1321	2	US-08-317-310A-64
12	209	4.3	408	1	US-07-609-716-65
13	209	4.3	408	4	US-08-475-411A-65
14	209	4.3	408	4	US-08-478-029A-65
15	205.5	4.2	1078	3	US-08-963-825-21
16	205.5	4.2	1078	4	US-09-570-573-21
17	205.5	4.2	1078	4	US-09-548-608-21
18	204.5	4.2	1060	3	US-08-931-820-3
19	204.5	4.2	1418	3	US-08-963-825-20
20	204.5	4.2	1418	4	US-09-010-999-1
21	204.5	4.2	1418	4	US-09-570-573-20
22	204.5	4.2	1418	4	US-09-548-608-20
23	203	4.1	960	4	US-09-219-849-6
24	202	4.1	552	4	US-09-219-849-7
25	201	4.1	720	4	US-09-219-849-4
26	201	4.1	777	1	US-08-642-255-53
27	200.5	4.1	633	1	US-08-642-255-73

Sequence 4, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 3, Appl  
Sequence 52, Appl  
Sequence 3, Appl  
Sequence 120, Appl  
Sequence 31, Appl  
Sequence 19, Appl  
Sequence 19, Appl  
Sequence 19, Appl  
Sequence 2, Appl  
Sequence 5, Appl  
Sequence 114, Appl  
Sequence 26, Appl  
Sequence 84, Appl  
Sequence 14, Appl  
Sequence 49, Appl

#### ALIGNMENTS

RESULT 1  
US-08-918-914-4  
; Sequence 4, Application US/08918914  
; Patent No. 5876963  
; GENERAL INFORMATION:  
; APPLICANT: Mitchell, Peter  
; APPLICANT: Hutchinson, Nancy  
; APPLICANT: Lawton, Michael  
; APPLICANT: Magna, Holly  
; APPLICANT: Yocum, Sue  
; APPLICANT: Murry, Lynn E.  
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/918,914  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0369  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 788 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1070094  
; US-08-918-914-4

```

; Patent No. 5773249
;
; GENERAL INFORMATION:
;
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
;
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
;
; INFORMATION FOR SEQ ID NO: 62:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1064 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-08-642-255-62

```

```
Query Match      4.7%; Score 232.5; DB 1; Length 1064;
Best Local Similarity 23.4%; Pred. No. 1.2e-07;
Matches 206; Conservative 32; Mismatches 324; Indels 309; Gaps 46;
```

QY	42	GPEGVGMWVOWAKSCSPCGVGVRISRTCOLTVQLHPSLPLP-----PRPPRHP 92
DB	49	GPXGAGPP-----AGPPG-----SRDPPG-----PGAGCPAGPGSGRDGDFGPPCAP 90
QY	93	EALLPRGO--GFRPQTSPETLIYRTQSRGRGF--LRGPASHLGREETAIRAAARRSL 148
DB	91	GPAGPPGSKGDGPPGAPG--PAGPPGSGRDGPPGAPGAPGG-----SR- 135
QY	149	RDPITKFMFGYGRVPFALPLHRNRHRSPRSLSLSRGEELIP-SPTPRAEPFSAN 207
DB	136	GDPGPPGAPG-----PAGEPP-----GSRDDGPPGAPGAPGPPGSGRG 172
QY	208	-----GSPTQTELPTELTSVHTPSPQAPELSPETAQTEVAPTRRAPLRHHRAQASGTPE 262
DB	173	DGP2PAGAGPAGPGSGRDGPPGAPGAPGPPGSGRDGPPGAPGPA--GPPSGRDGPP 230
QY	263	PSPTHSLGEGFFRASPOPRRSSQGWSAQVAGRBDP-----FPSVPRG-----R 309
DB	231	PGAPGAPGAPPG--SRGDPPG--PGAPGAPG--PGSGRDGPPGCAHGAPGPKAHGAPGPK 285
QY	310	GOQG--QGFWGWTGGTTHGPRLEP-----DQHPHGAWLLSNGPHASSLWSLFAPS 358
DB	286	GAHGAPGAPKGANG--PAGPKAGPAGPPGSGRDGPPGGA-----P-----CPA 327
QY	359	SPIPRCSGESQLRACSAQCPCPPQPDPRALQCAAFNSQEFMWGLYQWFHFTEVQGSORC 418
DB	328	GP-PGSGRD-----EGPPCAGAPGAG-----GPPCAGAPGAG----- 346

RESULT 2  
US-08-642-255-62  
; Sequence 62, Application US/08642255

Query Match	5.8%;	Score 284;	DB 2;	Length 788;
Best Local Similarity	20.2%;	Pred. No. 3.4e-11;		
Matches 189;	Conservative	76;	Mismatches 296;	Indels 370; Gaps 43;
QY	15	LLSLPQLCQDELVLSGHSLQTPEEG-----QGEGV--WGPVTQMASC-	56	
Ddb	115	IIVMLELVDCHQFDAH---PPLEEDGLASSTDKYDRKKRQGDDEAVGVGDMTWSHCS	171	
QY	57	-----SQPCGVGVQRSTTCOLPVQLHPSLPLP-----PRPRRIUEALLPRGGQPR	103	
Ddb	172	SNGHEVRSQACEYGRKTRRGCC--PARSAPQRVPAPPAQQAYAPRAPEYSACQQQQQREQ	229	
QY	104	POTSPETLPYRTQSRGRGGLRGPHASHLGRETQIBRAARSRLRDIPKGMFCYIGRVP	163	
Ddb	230	QQ-----REQQH-----REHQARLQHQOQQOQQOQQOQQRP-----QP	264	
QY	164	FALPLHRNRHRRSPPRS-----ELSLISRGEEAIPSTPPRAEPFSANGSPQTELPPTE	218	
Ddb	265	QFQPOPPPQPPQPGQSFGSTHELHLQREQOQQOQQOQQOQQOQQOQQNFPQQTPQQTQ	324	
QY	219	LSVHTPSQAEPILSPETAQAEVARTRP-----APLRHHP-----RAQASGEPPS	264	
Ddb	325	FGSQIQLQSGVPPQHPHQOQQOQQOPELERSPLDCHAAQLQQRMYSQYRENFNORHPAR	384	
QY	265	PTHSLGEGGF---FRASPQPRRSSGWASPPQVAGRRDPDFPSVPRGRGQOQOGPWGTGG	321	
Ddb	385	PKADPCPGGCFAPVQAPQOERPTPPVLAPVINTATQPLPQ-----CNLG-----	427	
QY	322	TPHGRLPEPDPOHSGAN-----LPLLNSGHASSLSLAFSPSPICRS-----GESE	369	
Ddb	428	-PYTRYEPAPPPACDGCGVCYNPPVSVGVHDNSDKS-----TCSTCGDGAKS	477	
QY	370	QLRACSQAACPPEQDDPALOCAFNESQEMGOLYWNEFTFEVQSGRCNELNCRPRGRPF	429	
Ddb	478	RRECESTNNC-----QGADYETP-----	500	
QY	430	VVRHTEKVODGTLCPGAPDICVAGRCSPGCDIGLGSRRPD-----GCVCVGDDST	483	
Ddb	501	-----PCQTW-S-EWCWSITC-SASC-----GSGORETRFCHGINRCBGDYE	542	
QY	484	CLRVSGNLTDREGPLGYOKILINTPAGALRLQIAQLRPSNYIALRGPGGRSIINGNAVD	543	
Ddb	543	SEQCS-----AGPCP-EWSQWEDWGQCSTCG-----GVAVR---QRTCLG-----	581	
QY	544	PGSVRAGTGTVRYNRPREEGGESLSARGPTTOPVDVMIFQENPGVFQYVISPP	603	
Ddb	582	-----VF-----GDHL-CGKPXTE-----QACDGGP	602	
QY	604	FILENPTPEPVPOLQEPIILVEPPELAPARPARTGTLQORVRIQWPMPAPPHPPTLGS	663	
Ddb	603	CSLMSPMOE-----	611	
QY	664	PAAYMKRVGHSACSASCKGKWVRPIFLCISRSGEELDERSCAAGA---RPPASPEPCHG	720	
Ddb	612	----W-----STCASCSGSKMR-----ROVCQFGTDCQGFNEESQFYCG	648	
QY	721	TPCPPYPWAGEWTSCRSQPGTQHROLQCRBFGGGSSVPPERCGHLPRPNI-TQSCQ	779	
Ddb	649	PPCAEWTEWCSEWSCSKCGPGQRTTRGC-----LGNPG-----QZATTCCGPSIETLCE	700	
QY	780	LRLCGHNEVSPWSQCSVRCGRQ-----	812	
Ddb	701	GOSCCNWSEWNMCBKECGGGQVRYIEYMFRTGCEWSFCSTQACEVGVQSRRCQCVG	760	
QY	813	NNG-----DEVSEQ-----ECASGPPFPQP	831	
Ddb	761	ESGHCICIGLAESQOCRLGTQC---PPKPP	787	

## RESULT

QY 419 ELNCRPRGFRVYRHTKVKQDGTLCQAGADICVAGRCLSPGCDGILSGRRRDPGCVG 478  
 Db 347 -----PGSR-----GDCPPAG-----GPAGPPSGRDPGPPGAPGAG--- 382  
 QY 479 GDDSTCLVSNLDRG-----GGLGQKILWIPAGALRLQIAQLRPSSNYLALRGPGGR 533  
 Db 383 -----PPSGRDPGPPGAPGAG-----PPGS-----RGDPGPPGAPGAP 419  
 QY 534 SIINGNNAVDPPGSRAGTVFRYRPPREBKGESLSAEGTTPQVDVYVIMFOEENPV 593  
 Db 420 PGRSGD--PGPPGAPGAG-----PPSGRDPGPPGAPGAGPP-----GSRGDPG 464  
 QY 594 FYQYVSSPPILNPPPEPPVQLOPELLVPEPLAPAPARTCTLQORVIRPOMPA 653  
 Db 465 PGAGPAGPPGSRGDPGPPGAPGAGPPGSRGDPGPPGAPGAGPPGSRGDPGPPG 517  
 QY 654 PPHRTPLGSPAAYKRWKRGHSASCSGKGYWRPFLICISRESGEELDERSCAAGARPPA 713  
 Db 518 PPAHGPAGPKGAHGPAGPKGAHGPAGPKGAHGP-----AGPK-----GAPGPA 561  
 QY 714 SPECHGTGPPYWEAGTSCSRSCGPGTQHRQLQCRQEBGGGSSVPPBERCGH----- 768  
 Db 562 GPPGSRGDPGPPGAPGAGPPGSRG--DPGPP-----GAPGAPGPPGSRGDPGPPG 610  
 QY 769 LPRNITQSCRLCLGHWEGVSPWQSVRCGRGQRQSRQVRCVGNNGDEVSEQCASGPP 828  
 Db 611 APGP-----AGPP-----GSRGD--PGPPGAPGPA 633  
 QY 829 QPPSREACDMGPTTANFHSWSSKVSPEPP 859  
 Db 634 GPPGSRG--DPGPPGAPGAGPPGSRGDPGPP 663

RESULT 3  
 US-09-219-849-5  
 ; Sequence 5, Application US/09219849  
 ; Patent No. 6150081  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN HEERDE, GEORGE V.  
 ; APPLICANT: VAN RIJN, ALEXIS C.  
 ; APPLICANT: BOWSTRA, JAN B.  
 ; APPLICANT: DE WOLE, FREDERIK A.  
 ; APPLICANT: MOORBEEK, ANDREAS  
 ; APPLICANT: WERTEN, MARC W.T.  
 ; APPLICANT: WIND, RICHEL D.  
 ; APPLICANT: VAN DEN BOSCH, TANJA J.  
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN  
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
 ; FILE REFERENCE: 2728-2  
 ; CURRENT APPLICATION NUMBER: US/09/219,849  
 ; CURRENT FILING DATE: 1998-12-23  
 ; NUMBER OF SEQ ID NOS: 50  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 960  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Illustrative  
 ; OTHER INFORMATION: amino acid sequence  
 US-09-219-849-5

Query Match 4.6%; Score 224.5; DB 4; Length 960;  
 Best Local Similarity 25.2%; Pred. No. 3.7e-07;  
 Matches 194; Conservative 25; Mismatches 323; Indels 227; Gaps 41;  
 QY 36 PTEEGGPGGVWGVWQVASCQGVQVRRSRTCLQTLVQLHPSLPLPPRRPHREAL 95  
 Db 95 PGSRDPGPPGAGP-----AGPG-----SRDPGPPGAP--GAPGPPGSRDPGPPGAP 140  
 QY 96 LPRG-----QGRPQTSPTLPLRYTQSRGGRG--LRGFASHIGRETOEIRAARRSLRD 150

Db 141 GPAGPPGSRDPGPPGAPG-----PAGPPGSRDPGPPGAPGAPGPPGS-----RD 184  
 QY 151 PIKPMFGYGRVPFALPLHNRHRRHPSRPSRSELSSRGEAEIAPSPTPRAEFPSSAN--GS 209  
 Db 185 PGPPGAPG-----PAGPP-----GSRDPGPPGAPGAPGPPGSRDPG 221  
 QY 210 PQTELPTELSVHTPS--PQAE--PLSPETAQTEVAPTRPAPLRHHPRAQASGTEPSPST 266  
 Db 222 PGAHGPAGPKGAHGPAGPKGAHGPAGPKGAHGPAGPKGAPG-----AGPPGSRDPGPP 275  
 QY 267 HSLGEGGFFRASPQRRPSSQGWASQVAGRRRDPDPFSPVPRGRGQQG---GFWGTGGTP 323  
 Db 276 GAGPAG--PPGSRDPGPPGAPGAPGPP--GSRDPG--PGAPGAPGPPGSRDPGPPGAPG--P 331  
 QY 324 HGRPLRDPDHPQAWLPLLSNGPHASSLSLAPSP--IPRCSGESEQLRACSAQAPPE 382  
 Db 332 AGPPGSRDPGPPGAPGPA---GPPGSR-----DPGPPGAPGAPGPPG-----SRDPGPPG 378  
 QY 383 QDPRALQCAAFNSQEFMGQLYQWEPFTEVQGSQRCCLNCRPRGFRFYVRHTEKVDGTL 442  
 Db 379 APGAG-----PPGSRDPGPPGAPGAPGPPGSR-----DPG-- 409  
 QY 443 CQGAPDICVAGCLSPGCDGILG-----SGRR-----PDGCGVCGDDSTCLVSGNL 491  
 Db 410 -PPGAPGAPGPPGSRDPGPPGAPGPPGSRDPGPPGAPGPPGPPGPPGPPGPPGPPG 468  
 QY 492 TDGR--GPLGYQKILWIPAGALRLQIAQLRPSSNYLALRGPPGSRSIINGNNAVDPPGSR 549  
 Db 469 GPKGAHGPAG-----PKGA-----HGPAGPKGAH-----PAGPKG 499  
 QY 550 AGGTVFRYRPPREBKGESLSAEGTTPQVDVYVIMFOEENPVFYQYVSSPPPILENP 609  
 Db 500 APGPA-----GPPGSRDPGPP--GAPGAPGPP-----GSRDPGPPGAPGAPGPPGSRDPG 547  
 QY 510 TPEPPVQLOPEILLRVEPPLAP-----APRPARTGTLQROVRIPOMPAP 654  
 Db 548 PGAGPAGPPGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGPPG-----RDPGPPGA 602  
 QY 555 PPHRTPLGSPAAYKRWKRGHSASCSGKGYWRPFLICISRESGEELDERSCAAGARPPAS 714  
 Db 603 PGAPGPPGS-----RDPGPPGAPGAPGPPG-----SRDPG-----PPGAPGAPG 641  
 QY 715 PEPCHGTGPPYWEAGTSCSRSCGPGTQHRQLQCRQEBGGGSSVPP 763  
 Db 642 P-PGSRDPGPPGAPGAPGPPGSRDPGPP-----GAPGAPGPP 677

RESULT 4  
 US-09-041-886-23  
 ; Sequence 23, Application US/09041886  
 ; Patent No. 6233872  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bredesen, Dale E.  
 ; APPLICANT: Rabizadeh, Sharroz  
 ; TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
 ; TITLE OF INVENTION: Polypeptides and Methods of Use  
 ; NUMBER OF SEQUENCES: 72  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Campbell & Flores LLP  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 92122  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/041,886  
 ; FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: 8-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-041-886-23

Query Match 4.6%; Score 224.5; DB 4; Length 1185;  
Best Local Similarity 20.8%; Pred. No. 4.8e-07;  
Matches 212; Conservative 50; Mismatches 293; Indels 465; Gaps 52;  
QY 16 LSLPOLCDEVLGSHLSQTETEGQGEVWGPVWVWASQPCQGVGVQRRST----- 70  
DB 96 LRPQSPDLSLGRSL---NDGSG-----SDPRDIDQDNSTSPSIYS 137  
QY 71 -----COLTVQLHPSLPLPRP-----PRHPEA-----LLPRG----- 99  
DB 138 PGSVENDSSSGLSQGPARYHPPLPFPSPQPPDSTPROPEASFEPHPSVTGTGHAP 197  
QY 100 -----QGP--RPQTSPTLPLRYTQSRGGGGLPGPASHLGRRETOEIRARRSL 148  
DB 198 MEPTSRMFQAPGAPPHPQLYP-----GGTGGVLSGP----- 231  
QY 149 RDIKPMFGYGRVFPALPHNRHRRSPRSELSLSSRGEAIPSPTPRAEPFSANG 208  
DB 232 --PMGPK--GGGAASSVGGPNGKQHP--PPTTPTIS--VSSGASGAPPTKPTTTPVGGN 284  
QY 209 SPQT-----ELPP-----TELSTVTPSQAE-----PL 231  
DB 285 LPSAPPANFPHVTNLPPPPALRNLNNAASPPGLGQAFLGHLPSYAMGQMGGLP 344  
QY 232 SPETAQTEVAP-----RTRPAPLRHHPRAQASGTEP----- 262  
DB 345 GPEKGPT-LAPSPHSLPAPSSAPAPPMRFPYSSSSSSSSSSSSSSSSSSSSSSSPFPASQ 403  
QY 263 --PSPTHSLGEGFFRASPOPRR-----PS-----SQGWASQVAGR-----RPDPFP 303  
DB 404 ALPSYHSPPTPTSLSVSNQPPKYQTPSLPSQAVWSQGPPTPPYGRLLANSNAHPGFP 463  
QY 304 -----SVPRGRGQOQGPWGTGTPHGRPLEDP-----Q 333  
DB 464 PSTGAQSTAHPPVSTHHHHHQQOQQOQQOQQOHHGNSGPP-PPGATPHLEGSSSHA 522  
QY 334 HGCAMLPILSN-----GPHASLWLEFAPSPIPRCSESEQLRACSOACPP 381  
DB 523 HPYAMSPSLGSLRPPGPAHLPPHPSQVSYQAQNG--PPVSSSSSSSSSTSGSYPC 580  
QY 382 EQDPDRALQCAAFNSOEFMGLYQWEPFTEVQSOR-----CELNCRPRGFRFVVRHTE 435  
DB 581 SHSP-----SQPGQAGYPPPPVPTVTSSATLSTVIATVASSPAGYK----- 624  
QY 436 KVQDGLQOPGAPDICVAGRCLSPGC-DGILSGRRPDCGCGGDDSTCLRLVSNLDR 494  
DB 625 -----TASPPGPPPP--YGRAPSPGAYKATPPGYK-----GSPSEFR-----T 662  
QY 495 GGPLGYQKILWIPAGALRLQIALRPSS-----NYIALRPGGGRSLINGNAVDPPGYSR 549  
DB 663 GTPPYRGIS-PPAGP-----GTFKPGSTVGVGPLPPAGPSGLPSL-----PP----- 705  
QY 550 AGGTVFYRNRPREEKGESLSLGGTTTQPDVYIMFQENGVYQYVSSPPFLENP 609  
DB 706 -----PPAAPASGPPLSATQIKQPAEY-----ETP-----ESPVPAPKSP 742

QY 610 TPEPPVPOL----- 618  
DB 743 SPPPKVVDVPSHASQSAREFNKHLDRGFNSCARSDLYFVPLEGSKLAKKADLVEKVRREA 802  
QY 619 -----QPEILR-----VE-DPLAPAP-RPARTPG- 640  
DB 803 EQAREEKEREREREKEREREKELERSVKLAQEGRAPVECSLGVPHRPPPEPS 862  
QY 641 -----TLQROVRIFOMPAP---PHP-RTPLGs-----PAAWK 669  
DB 863 AVATVPPYLPDTPALRTLSEYAR-PHVMSPGNRNHPFVYPLCAVDPLGLGYNVPALY-- 919  
QY 670 RVGHSACSACGKGVWRPIFLCISRESGELDERSCAAGARP-----PASPCHGTCP 724  
DB 920 -----SSDPA-----ARERERERDRDLRDKPGFEVFKPSELEPLHGVGP 961

RESULT 5  
US-09-413-814-78  
Sequence 78, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
TITLE OF INVENTION: heteropolyketide compounds  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 78  
LENGTH: 882  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-78

Query Match 4.5%; Score 220; DB 4; Length 882;  
Best Local Similarity 22.8%; Pred. No. 6.7e-07;  
Matches 165; Conservative 46; Mismatches 240; Indels 274; Gaps 35;

QY 29 SGHSLQTPTEGQGEVW--GPVWVWASQSQ----PCGVGVQRRSRTCOLPTVOLHPSLP 83  
DB 281 STHASEPASLAEALGOLWEAGHAVDTARPRGRPPARSACRTRSSRGTSRRAAAP 340  
QY 84 LPP--RPPRHPALLPRGQPPRQTSPTETILYETQSRGGPLRGA-----SHLGR 134  
DB 341 PPPSRRTPTFSR--PRTPSRRRSPARTPGPSRRRPSARSAPAIWERRPRRDR 398  
QY 135 EETGEIARAARSRLDKPKMGFGYGRVPPFALPLHRRHRPRSPRSELSSISRGEEAI 194  
DB 399 RPRLLRPARLARDP-----GDVPD--PRAARRAPAA----- 431  
QY 195 PSPTFRAEPFSANGSPOTELPPTLSTVHTFSPQAEPLSPETAQTEVAPRTPAPLRHHR 254  
DB 432 -GALPGADGRCARGGRRRARRGAAARVPPDRADP--PRRPAPVY-RAAPAVVRSAR 487  
QY 255 AAGSTEPSPTSHISLGGEGFFRASQPPRRSSQGWASQ-----VAGRR--- 298  
DB 488 AQQGVQHPVKGADGPARRRRAPQPVDR-----AAPRGAADDVQGRRAVPRRARG 543



825 PGOIAGQGRVGLPQGRGERGPPGLPGP-----SGEPKQ--GPSGASGERGPPGPMG 877  
718 CHGTPCPPYWEAGWT-----SCSRSCGPGTQHRQLQCRQEFGGGSSVPPERCGLHLP 772  
878 PPLAGPPEGSGREGAPGAEGSPGRDGSPP-----AKGDRGETGPAGPPGAXGAXGAP 930  
773 N-ITQSCQLRLCGHWEVGSVMSQSVRCGRQSRQVRCVGNNGD--EVSEQ-----821  
931 GPVGPAGKSGDRGETGPAGPAGPVGVPAGAGPAGPQ-----GPRDKGETGEGDGRGKXG 986  
822 ---ECASGPPPPSREACDMGPTTAMFHSWSSKVSPEPP 859  
987 RGFSGLQPPGPPGSPG-EQGP-----SGASGAPGPRGP 1019

RESULT 7  
US-09-570-573-18  
Sequence 18, Application US/09570573  
Patent No. 6342361  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570.573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1341 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (I)  
US-09-570-573-18

Query Match 4.4%; Score 216.5; DB 4; Length 1341;  
Best Local Similarity 22.9%; Pred. No. 1.9e-06;  
Matches 243; Conservative 62; Mismatches 411; Indels 345; Gaps 56;

QY 22 CLDQEVLSGH-----SLQTPTEEG-----QGPEGVGVWVQWASCSQCGVGVQRRSRT 70

81 CPGAEPVEGECPCVCPDGSSEPTDQTTTVEGPKGDTGPRGPRGPPGDRGDIPOQ---137  
71 CQLPTVOLHPSLDLPPRPRHP-----EALLPRG- 99  
138 -----PGLPGPPGPPGPPGGLGNFANPOLYGVDEKSTGTGTSVPGMPSGPRGL 188  
100 QGPRPQTSPEL--PLYRTQSRGRGGL--RGPASHLGRE-ETQETRAARRRLRDIKP 154  
189 PGPPGAPGPGZGF7GPGZPCZPCASGPMGPRGPPGPKGBGBGZAGKPRGZRGPP--P 246  
155 GMFGYGRVFPALPLHNRHPRSPRSELSLSSRGEALPSPTPRAEPFSANGSPQTEL 214  
247 GPZGARGLPGTAGLPMKGH-----RGFSLBGAKE-BAGPA-GPKGZP-----GSPGZBG 295  
215 PPTLSVHTPSPO---AEPLSPET-----AOTEVAP-----RTRPAPLRHHHPRAQASCTE-- 261  
296 APGZMG---PPGPKNSGEGCAPGSKGDTGAKGEPGVGVQPPGPGAGEBGRGARGEPG 353  
262 -----PPSPHSLGEGGF-----FRASFPQPRPSSQGWASPOVAGR----- 297  
354 TGLPGPPGERGPGSRGFPDAGDVAGPKGPAERGSFPGAPK-----GSPGEAGRGEGAG 409  
298 -----RDPFFSVPRGRQOG-----QGFPGTGGTP 323  
410 LPNKGILTSFGSPGDPGKTGPPGAPQDGRPGP-PPGPGARGAGVMGPPGPKGAAGEP 468  
324 --HGPRLEPPDHPGAWLPLLSNGPHASSLSWLFAPSSPIPRCSGSEQLRACS----- 375  
469 GKAGERGVPGP--PGAVGPAGKDEGAQ-----GPPGPAGPAGERGEOGPAGSPGFQGL 521  
376 -----QAPCPPEQDPRAL-----QCAAFNSOEFNGQLYQWEPETEVOGSORCELNCR 423  
522 PGAPGPEAGKQEGQVPGDLGAPGSGARGERGFGF-----RGVGGPPGAP 571  
424 PRGFRFYVRHTEKVQDGTLCQPGAPDLCVAGRCLSPSCDGLG-----SGRRPDG 473  
572 PRG-----ANGAPNDGAKGDAGAPG--APGSOGAPQLQMGPERGAAGLPKGDGRGD- 623  
474 CGVCGGDDSTCLVSGNLTDRGGPLGYQKILWIPAGAL--RLQIALRPPSSNVLALRGP 531  
624 APKGDGSGPKDGVRLTGPIGPPG-----PAGAPGKSGSGSPGAPGTGARGAPG 676  
532 GRSTINGNWAVDPPGSYRAGGTVFYRNPREEG-----KGESLSA-----EGPT 576  
677 DRG-----EPGPPGAPGAG-----PPGADQPGAKGEPGDAGAKGDAGPPGAPGA 723  
577 TOPVDVIMFQENPGVYQYVVISPPPLENTPTE---PPVPQLQPEILRVEPPLAPA 632  
724 GPPGPIGNV---GAPGAKGARGSPGATGFGAAGRVPVPPGFS-----GNAGPPGPPG 775  
633 PR-----PARTGTLQORVRIQMPAPPHPRTPLGSPAAYWKRVHGSACSAS 679  
776 PAGKEGKGKPRGETCPAGRPCEV-----GPPGPPGAPGKSGPGA-----DGPAGAPGT 824  
680 CG-----KGWVRPFLCISRESGEELDERSCAAGARPPASPEP 717  
825 PPGIAGQGRVGLPQGRGERGPPGLPGP-----SGEPKQ--GPSGASGERGPPGPMG 877  
718 CHGTPCPPYWEAGWT-----SCSRSCGPGTQHRQLQCRQEFGGGSSVPPERCGLHLP 772  
878 PPLAGPPEGSGREGAPGAEGSPGRDGSPP-----AKGDRGETGPAGPPGAXGAXGAP 930  
773 N-ITQSCQLRLCGHWEVGSVMSQSVRCGRQSRQVRCVGNNGD--EVSEQ-----821  
931 GPVGPAGKSGDRGETGPAGPAGPVGVPAGAGPAGPQ-----GPRDKGETGEGDGRGKXG 986  
822 ---ECASGPPPPSREACDMGPTTAMFHSWSSKVSPEPP 859  
987 RGFSGLQPPGPPGSPG-EQGP-----SGASGAPGPRGP 1019

RESULT 8

US-09-548-608-18

; Sequence 18, Application US/09548608  
; Patent No. 6355442  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/548,608  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1341 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (I)  
US-09-548-608-18

Query Match 4.4%; Score 216.5; DB 4; Length 1341;

Best Local Similarity 22.9%; Pred. No. 1.9e-06;

Matches 243; Conservative 62; Mismatches 411; Indels 345; Gaps 56;

QY 22 CLDQEVLSGH-----SLQTPTEEG-----QGPEGVGWVQVWASCSQPCGVGVQRRSRT 70  
DB 81 CPAEVEGECPCVCDGSESFDQETTVGEGKGTGPRGRGAGPRGRDGIQO--- 137  
QY 71 COLPTVOLHPSLPLPRPPRHP-----PGLGPPGPPGPPGLGGNFAPQLYGYDEKSTGTSVFPMPGSPGRGL 188  
DB 138 -----PGLGPPGPPGPPGPPGLGGNFAPQLYGYDEKSTGTSVFPMPGSPGRGL 188  
QY 100 QGPRQTSPEL--PLYRQSRGCGPL--RGPAHLGRE--EQEIRAARRSLRDPKIP 154  
DB 189 PGPGAPGPGGFGPGCGZPCZPGASGPMGPPGPPGKBGBGZAGKPGRPGZRP--P 246  
QY 155 GMFGYGVFPFALPHNRHRRSPRSELSLSSRGEAIPSPTPRAEFPSSANGSPQTEL 214  
DB 247 GPZGARGLPCTAGLPKMGH-----RGFSLBCAKG-BAGPA-GPKGZP-----GSPCZBG 295  
QY 215 PPTELSVHTPSQ---AEPLSPET-----AOTEVAP---RTRPAFLRHHPRAQASCTE-- 261

DB 296 APGZMG--PPGPKNSGEPGAPGSKGDTGAKGEPGVGVQPPPGPAGEGKRGARGEPGP 353  
QY 262 -----PPSPHSLGEGGP-----FRASQOPRRPSSQGWASPOVAGR----- 297  
DB 354 TCLPGPPGERGGPGSRFPFGADGVAGPKPAGERGSPGAPK-----GSPGEARPGEAG 409  
QY 298 -----RPDPFSPVRCRGOOG-----OGPMGTGTGP 323  
DB 410 LPGAKGLTGSCGSPGPGDKTGPFGAGODGRPG--PGPPGARGAGVGMFPFGKAGAPG 468  
QY 334 --HGPRLEPDPQHPGAMLPLLNSNGPHASSLSLFPAPSPIPRCSGSESEQLRACS----- 375  
DB 469 GKAGERGVGP--PGAVGPAGKDGGAQA-----GPPGAPGAPAGERGEGQSPAGSPGQGL 521  
QY 376 -----QAPCPPEQDPRAL-----QCAAFNSOFFMQLYQWEPFTEVQCSQRCCLNCR 423  
DB 522 PGPAGPGEAGKPGEGQGVFGDLGAPGSGARGERGEGE-----RCVQPPPGPAG 571  
QY 424 PRGFRFYVRHTEKVDGTLQCPGAPDICVAGRCILSPGCDGILG-----SGRRPDG 473  
DB 572 PRG-----ANGAPNDGAKGDAGAPG--APGSQAGPLQGMPPGERGAAGLPKPGDRGD- 623  
QY 474 CGVCGDDSTCRLYSGNLTDRGGPLGYQKILWIIPAGAL--RLQIAQLRPSNLYLALRPG 531  
DB 624 AGPKGADGSPGKGVRLTGTGIPPG-----PAGAPGDKGESGSPGPGAGTGAAGAPG 676  
QY 532 GRSIINGNVAVDPPGSGYRAGGTVERYNRPREEG-----KGESLSA-----EGPT 576  
DB 677 DRG-----EPGPPGAPGAG-----PPCADGPGAKGPGDAGAKGDAGPSPGAPGA 723  
QY 577 TOPVDVYMFQENPGVYQVVISPPPILENPIPE-----PPVQLQPEILRVPPPLAPA 632  
DB 724 GPPGPIGNV--GAPKAGKARGSAGPGCATGFPGAAGRVGPPGPS-----GNAGPPGPPG 775  
QY 633 PR-----PARTGTLOQRVIRPQMAPPHPTPLGSPAPAAWKVHVSACSAS 679  
DB 776 PAGKEGKGRGTGAPRGEV-----GPPGPPGAPGKGSFGA---DGPAGAPT 824  
QY 680 CG-----KGWVRPIFLCISRESGSELDERSCAACARPPASPPEP 717  
DB 825 PPGQIAGQGVVGLPQGRGERGFPGLPFP-----SGEPGKQ--GPSGASGERGPPGPMG 877  
QY 718 CHGTCPPPYWEAGWT-----SCSRSCGPGTQHRQLQCRQEFGGGSSVPPERGCHLPRP 772  
DB 878 PPGLAGPPGESGREGAPCAESPGDGSFG-----AKDRGTGTGAPGPGAXGAXGAP 930  
QY 773 N-ITQSCQLRLCGHWEGSPWSQSVRCGRGSRQVRCVGNNGD--EVSEQ----- 821  
DB 931 GPVGPAGKSGDRGTGPAGPAGVPVGPAGARGPAGPQ-----GPRGDKGTGEGQDRGKNGH 986  
QY 822 ---ECASGPPPPPSREACDMGPCITAFHSDMSKVSPEPP 859  
DB 987 RGFSLQGPFGPPGSPG-EQGP-----SGASGAPGPRGP 1019

RESULT 9

US-08-642-255-72  
; Sequence 72, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/642,255  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: ROWLAND, Bertram I.  
 REGISTRATION NUMBER: 20,015  
 REFERENCE/DOCKET NUMBER: A5556-3/BIR  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 494-8700  
 TELEFAX: (415) 494-8771  
 TELEX: 910 277299 FHT UR  
 INFORMATION FOR SEQ ID NO: 72:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1065 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-642-255-72

Query Match 4.4%; Score 213; DB 1; Length 1065;  
 Best Local Similarity 23.7%; Pred. No. 2.4e-06;  
 Matches 208; Conservative 32; Mismatches 341; Indels 296; Gaps 46;

QY 62 VGVQRS---RTCLPTVQLHPSLPLPRPRHPEALLPRG---QGRPOTSPTETLPL 113  
 DB 4 VVLRDWDENPGVTQNLRLAAHPTFASDPMGAHGAG--PKGAHPAGPKGAQGAG-- 58  
 QY 114 YRTQSRGRGP--LRGPASHLGRTEQETRAARRRLDPIKPMFGYGRVPFALPHRN 171  
 DB 59 -PGSGRDPGPPGAQGPAGPGG-----SR-GDPGPPGAQ----- 91  
 QY 172 RHRPRPPRSELSISRGEEAIPSTPRAEPFANGSPQTELPTELTSVHTPSQAEPL 231  
 DB 92 ---PAGPG-----GSRGDPGPPGAQGPAGPGSGRDPG---PFGAQPAGPGSGRDP 138  
 QY 232 SPETACTEVAP---RTRAPLRHHPRAQSGTE---PPSPHSLGEGFFRASPQRRP 284  
 DB 139 GPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPG--SRGDPG--P 195  
 QY 285 SSQGWASPVAGRRPDPFVSFRG-----RQOQ---QGPWGTGTF----- 322  
 DB 196 GAQGPAGP--GSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQ 253  
 QY 323 ---PHGPRLEDPQHPGAWLPLLSNGPHA--SSLSLIFAPSSPIPRCSGESEQLRACSOAP 378  
 DB 254 PAGPGSGRDPG--PGAHGAPGKGAHPAGPKGAHPAGP--KGAHPAGPKGA--QGP 308  
 QY 379 CPP-----EQDPRALQCAAFNSQEFMGQLYQWEPFTEVQSORCEINCRPRGRFYVRHT 434  
 DB 309 AGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGR 333  
 QY 435 EKVDGTLQCPGAPDLCVAGRLS-----PGCDGILGSG-----RRPDGCGVCGG 479  
 DB 334 -----GDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGR 388  
 QY 480 DSTCLVSGNLTDRGCGLYQKILWIPAGALRLQIAQLRPSNVIALRGPGCRSLINGN 539  
 DB 389 DP-----GPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAGPGSGR 425  
 QY 540 WAVDPGSTRAGTVRYNRPREKGSLSAEGTTPQVDYIMIFQENPGVYQYVI 599  
 DB 426 --PGPPGAQGPAG-----PGSGRDPGPPGAQGPAGPG-----GSRGDPG----- 463  
 QY 600 SSPPTILENPT-----PEPPVPO--LQPEILRVERPLAPAPARTPTGTLQQRVIP 649  
 DB 464 ----PPGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQGPAG--CPGSGRDP----- 515

QY 550 QMAPPHPRTPLGSPAAYKRVGHSACSCGKGVWRPIFLCISRESGEBLDBSCAGA 709  
 DB 516 --PGPPGAHGPAGPKGAHPAGPKGAHPAGPKGAHPAG-----AGPK-----GA 557  
 QY 710 RPPASPEPCHGTFCPP---YWEAGEWTSCSRSCGPGTQHRQLQCRQEFGGGSSVPPERC 767  
 DB 558 QGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPPGAQ-----GPAGPGSGRDP----- 606  
 QY 768 HLPRLNITQSCQLRLCGHWEVGSQSCVRCGRGQSRQVRCVGNNGDEVSEBECASGP 827  
 DB 607 -----GPPGAQGPAGPG-----GSRGD-----PGP 626  
 QY 828 -----PQPPSREACDMGPTTANFHSWDSKVSPEPP 859  
 DB 627 PGAQGPAGPGSGRDPGPPGAQGPAGPGSGRDPGPP 663  
 RESULT 10  
 US-08-931-820-1  
 ; Sequence 1, Application US/08931820  
 ; Patent No. 6010863  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: Assay for collagen degradation  
 ; NUMBER OF SEQUENCES: 4  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/931,820  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: EP 96202596.1  
 ; FILING DATE:  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1057 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; TISSUE TYPE: Collagen type I  
 ; US-08-931-820-1

Query Match 4.3%; Score 210.5; DB 3; Length 1057;  
 Best Local Similarity 23.6%; Pred. No. 3.5e-06;  
 Matches 235; Conservative 49; Mismatches 382; Indels 329; Gaps 55;

QY 36 PTEEQ-----GPEGVWGPWQWASQPCGVQVRRSRTCLPTVQLHPSLPLPRPRRH 91  
 DB 146 PGERGRPCAPGARGNDGATGAAGPPGTG-----PAGEPGF 184  
 QY 92 PEALLPRG-----QGRPOTSPTETLPLYNTQSRGPG-----LRGPASHLGRTEQIRA 142  
 DB 185 PGAVKAGCEAGPQGRSGEPQGV-----RGEPPGPPGAGAGPAGNPGADQPGKAG 237  
 QY 143 ARSRRLRDPKPMFGYGRVPFALPLHNRHRRSP--PRSELSI-----SSRGEEAI 194  
 DB 238 ANGA-----PGIAGAPGFFGA-----RGSPGPGPGTGCARGLVGPAGSGSGS- 285  
 QY 195 PSPTPRAEPFANGSPQTELPTELTSVHTPSQAEPLSPETAQTEVAPRTRPAP--LRHH 252  
 DB 286 ----NKGEPPGA--GPQG--PP-----GPSGEGKRGPNCEAGSAGP---PGPPGLRGG 328



Qy	253	PRAQ----	ASGTBPPBETHSLGEGFFRASPPRRSSOGNASPPQVAGR-----	297
		:	:	
Db	329	PGSRGFGDGVAGPK----	GPAG-ERGSPPGAPGK-----GSPGAGRGPEAGLPGAKG 378	
		:	:	
Qy	298	-----	RDPFPPSVPRGRGQG-----QCPMTGTGTP--HGPR 327	
		:	:	
Db	379	LTGSPSGPGDGKTPPGPAGQDGRFCP-	PGPPCARGQAGVMGFCPPKGAAGEPPKAGER 437	
		:	:	
Qy	328	LEPDQHPHGAWLPLLSNGPHASSLSL	FAPSSIPKCSSESQLRACS----- 375	
		:	:	
Db	438	GVFCP--PGAVGAGKDGEAGAQ-----	GPPGPAGAGERGEOGPAGSPFGQGLPGAPG 490	
		:	:	
Qy	376	--QAPCPEOPDPRAL----	QCAAFNSOEFMGOLYQWEPFTEVOGSORCELCNCRPGFRF 429	
		:	:	
Db	491	PGEAGKGEQGVFGDLCAPGSCARGEPGE-	-----RGVQGGPPGAPRG----- 537	
		:	:	
Qy	430	YVRHTEKVDGTLQCPGADICVAGRCL	SPCDOGILG-----SGRRPDGCVCGG 479	
		:	:	
Db	538	--ANGAPNDGAKDAGAPG-	APGSOCAPGLQCMFPGERGAAGLPGKDGDRG-AGPKGA 592	
		:	:	
Qy	480	DDSTCLRVSNLDRGGLGVQKILWIP	AGAL--RLQIAQLRPSSNYLALRGPGRSIIN 537	
		:	:	
Db	593	DGSPGCKDVRGLTGPIGPPG-----	PAGAPDKGESGSPGAGTGAARGAPDRG--- 642	
		:	:	
Qy	538	GNWAVDPPGYSRAGGTVFRYNRPREEG-	-----KGESLSA-----EGTTPQVDV 592	
		:	:	
Db	643	--EPGPPGPAFG-----	PPGADQPGAKGEPGDAGAKDAGPPGAPGAPGPGPI 632	
		:	:	
Qy	593	YMIQTEENPGVYQVYVSSPPIL	ENTPE----PPVPQLQPELLRVEPPLAPR--- 634	
		:	:	
Db	693	GNV--GAPGAKGARGAGPPCATG	FPAGRVPFGPS-----GNAGPPGPPGAPGK 744	
		:	:	
Qy	635	-----PARTGTLQVRIFQMPAP	PHPTPLGSPAAAYWKRVGHSACSSCG----- 681	
		:	:	
Db	745	KGPRGETGPAGRPEV-----	GPPPPGPAGEKSPGA---DGPAGAPGTPGPGGI 793	
		:	:	
Qy	682	-----KGWRPIELCISRESGEEL	DERSCAAGARPPASPEPCHGTPC 723	
		:	:	
Db	794	AGQVRVGLPQGRGERFPGLFPG-----	SCEPKQ--GPGSASGERGPPGMPGCLAG 846	
		:	:	
Qy	724	PPYWEAGWT-----SCSRSCPGTQ	HRQLOCROEFGGGSSVPPERCGLHLPARN-ITQS 777	
		:	:	
Db	847	PPGESRGEGAPCAEGSPGRD	SPG-----AKDGRGETGPAGPGAPGAPGAPGPA 899	
		:	:	
Qy	778	COLRLCWHWEVGSWWSQCSVRC	GRQSRQVRCYVGNND--EVSEQ-----ECA 824	
		:	:	
Db	900	GKSGDRGTGPAGPAGVPV	PAGAKGPAGPQ-----GPRGDKGETGEQDGRGKTHRGFSL 955	
		:	:	
Qy	825	SGPPQPPSREACMDGPTTAW	FHSDWSKVSPEPP 859	
		:	:	
Db	956	OGFPFPPGSPG-EOGP-----	SGASGAPGRPG 982	
		:	:	

```

RESULT 11
US-08-317-310A-64
; Sequence 64, Application US/08317310A
; Patent No. 5958701
; GENERAL INFORMATION:
; APPLICANT: WHITE, Morris F.
; APPLICANT: SUN, Xiao Jian
; APPLICANT: PIERCE, Jacquelyn H.
; TITLE OF INVENTION: THE IRS FAMILY OF GENES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: ASCII text
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/317,310A
, FILING DATE: 03-OCT-1994
, CLASSIFICATION: 435
, ATTORNEY/AGENT INFORMATION:
, NAME: Louis Myers
, REGISTRATION NUMBER: 35,965
, REFERENCE/DOCKET NUMBER: JDP-022
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (617)227-7400
, TELEFAX: (617)227-5941
, INFORMATION FOR SEQ ID NO: 64:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1321 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
, US-08-317-310A-64

```

	Query Match	4.3%; Score 210.5; DB 2; Length 1321;
	Best local similarity	21.0%; Pred. No. 4.6e-06;
	Matches	196; Conservative 79; Mismatches 313; Indels 345; Gaps 48;
QY	59	PCGVGVRRSRITCOLPTVOLHP-----SLPLPPRPHREALLPRQGQ 101
DB	417	PAG-GALQHRSRMKPVHSPAAATPSGLSSSGHGSGSYPLPGSHPLHLPHHPQG 475
QY	102	PRPQT-----SPELPLYRTQSRGRGGPL 125
DB	476	QRPSGGSASAGSPDPGFMSLDYEGSSPDLRAFSSHRSNTPESIATTPARDGSGGEL 535
QY	136	RG-----PASHLGR-----EETQETRAARRSLRDPIKPMFGYGRVPFALPHNR 172
DB	536	GYGMSMDRLSCHGRPYRVSDGAQDLRLRKTYSLTPEAR-----QRQV 583
QY	173	RHRPPRPSELSLI-----SSRGEERAIPTTPRAEPFSANGSQOTELPTELTVH--- 222
DB	584	POPSSASLDEVTLMRAFTEGSSGRCLCPFPASSPKV--AYNPYPE-DYGDTIEIGHSKS 639
QY	223	-----TPSPAELPLSTPAOTE-----VAPTRTPAPLR-HIPRAQA-----SGT 260
DB	640	SSNLGADOGYXMTFGAALRSGGNPSCKSDDYMPNSPTSVSAPKQLQPLRALAAPPSGA 699
QY	261	EPPSTHSLGE-----GGFFRASQP RPSPSQ-----WASPOVAGRDPDPPFVPRGRG 310
DB	700	AVPAPPVGVRTFPYNGGKYKASSPAESPDSGYRWMCWSKLSENDDP-KLLPNGDY 758
QY	311	-QQOGPWGTGCTPHGRPLED-----POHGCAMLP LLNMGPHASSLWSL 354
DB	759	LNMSPSEAGTAGTP-----PDFSAALRGSGELKGIPGHICYSSLPR-----SY 801
QY	355	FAPSGPIPRCGESQOLRACSAQ-----CPEQDPDRALOCA-AFNISOEMGOLIOWEPT 410
DB	802	KAPCS-----CSDNDQYVLMSSPVGRILEERLEPQATEGATFGA---AGGSTQPHHS 854
QY	411	EVQGSOR-CELCNCRPGRFRFY---VRHTEKVQDG-TL-----CQGPAP----- 448
DB	855	AVPSMRPSAIGRPEGFLGORCAVRPTRLSLGTLQTLPSMQEYPLTPEKSPGEVINI 914
QY	449	DICVAGRCLSQCDGITLGSRRPDGCVGCCDDSTCLRIVSGNLTDRGGFLGTQKILWIIPA 508
DB	915	DFEGAETRLSPAPP LLSAAA-----SSSSLLSAS-----SFA 947
QY	509	GAL-----RLQIAQLRSSNYLALRGPGRSIIINGNAVDPPPGSYRAGGTVFRYN--- 558
DB	948	SSLGSGETPOTSSDRQSRLSDYMNLD FSPKS-----PKPTRSGDTVGSMDGLL 998
QY	559	-----RPPREGKGESUSAEGETTQPDVDYMI-----FOENPGV 593
DB	999	SPEASSYPPLPRPSTSPSSIQQLPUPA-PGDYLRLPPASAATSGPTAGSMSSEPQ 1057

QY 594 FVQY-----VSSPP-PILENTPPE-----PPV 615  
DB 1058 NGDYTEMAFGVAATPPQIVAPPKEGARVASPTSGLKRLSLMDQVSGVEAFLOVSQPD 1117  
QY 616 POLQPEILRVEPPLAPAPARTPCTLQVRIPQMPAPPHPRTPLGSPAPYMKRVGHS 675  
DB 1118 PHRGAKVIRADP---QGRRRHSETFSTTV-----TPV-SPSFAHNSKRHS 1163  
QY 676 CSASCGKWPRIFLICISREGEELDRSCAAGARPASP---EPCHGTFCPPPYWEAGEW 732  
DB 1164 ASV-----ENVSLRKSSEGSTL-----GGDEPPTSPGOAQL--VAVPPVFOARPW 1209  
QY 733 TSCSRSCGPGTQHQLOCRQEFGGGSSVPPER 765  
DB 1210 -----NPGQCALIGC-----PGSSSPMR 1230

RESULT 12  
US-07-609-716-65  
; Sequence 65, Application US/07609716  
; Patent No. 5514581  
; GENERAL INFORMATION:  
; APPLICANT: Ferrari, Franco A.  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: Functional Recombinantly Prepared  
; TITLE OF INVENTION: Synthetic Protein Polymer  
; NUMBER OF SEQUENCES: 118  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/609/716  
; FILING DATE: 06-NOV-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Berttram I  
; REGISTRATION NUMBER: 20015  
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 408 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-609-716-65

Query Match 4.3%; Score 209; DB 1; Length 408;  
Best Local Similarity 23.1%; Pred. No. 1.5e-06;  
Matches 145; Conservative 14; Mismatches 200; Indels 270; Gaps 29;

QY 62 VGVORR-----RCQLPTVOLHPSL-----PLPPEPRHPEALLPRGGGPRPQTS 107  
DB 4 VVLQRRDWNPGVYQLNRLAAHPFASDPMGACPGPGPGPGPGAPGPGP-PGPPGP 62  
QY 108 PETPLRYTQSRGGGRLGPASHLGRETOEIRAAARRSLRDP IKPGMFGYGRVFPALP 167  
DB 63 P-----GPGFVGSFGA-----PGPPGPGP----- 82

QY 168 LHRNRHRSPRPSSELSLSSRGEEAIPSTPRAEPFSANGSPQTELPTELTVSHVTPSQ 227  
DB 83 -----PGPPP-----CAPGPPGPGPPGPPGAGPVGSPGAPGPP-----GPPGPP 123  
QY 228 AEPLESETAQTEVAPTRPAPLRRHHPRAQASGTEPSPSTHSLGEGGFFRASQPPRESSQ 287  
DB 124 GPPCAP-----GPPGPPGP-----PGPPGAGPVGSPG-----APGP--PGPP 159  
QY 288 GWASPOVAGRRRPPFFSVPRGQQG-QGFWGTGTPHGRLEPDPQHFGAWLPLLSNGP 346  
DB 160 GPPGPGCA---PGP-PCPPGPPGPPGAGPVGSPGAPGPPGPPGPPGPPGPPGPP 215  
QY 347 -----HASSLWSLFASSPPIPRCSGESEQLRACSAQCPPEQPDPRALQCAAFNSQEEFG 401  
DB 216 PGPPGAGPVGSPGAPGPPG-----PGPPGPPGAPGPPG----- 250  
QY 402 QLYQWEPFTEVQSGQRCELNCRPRFRFYVRHTEKVQDGTLCQPGAPDICVAGRCILSPGC 461  
DB 251 -----PPG-----PPGPPG--PAGFVGSFGA 369  
QY 462 DGLSGRRRDPGCGVGGDDSTCLVSGNLTDRGGLGYQKILWIPAGALRLQIALRPS 521  
DB 270 PGPPGPPGPPGPPGAPGPP-----GPPG-----PPG----- 296  
QY 522 SNYLALRGPGRSIINGNWAVDPPGSYRAGTVRYNRPRREGKGESLSAEGTTQPVD 581  
DB 297 -----PGPAGVGSFGAPGPPG-----PPGPPGPPGAPGPPGPPGPP-- 333  
QY 582 YMIFQENFGVYQYVSSPP-----PILENTPTEPPVQLOPEILRVEPPLAPAPAR 637  
DB 334 -----GPPGAGPVGSPGAPGPPGPPGPP-----GPPGAPGPP----- 365  
QY 638 TPGTLQVRIPQMPAPPHPRTPLGSPAA 666  
DB 366 PGP-----PGPPGPPGAGPVGSPGA 386

RESULT 13  
US-08-475-411A-65  
; Sequence 65, Application US/08475411A  
; Patent No. 6140072  
; GENERAL INFORMATION:  
; APPLICANT: Ferrari, Franco A.  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: Functional Recombinantly Prepared  
; TITLE OF INVENTION: Synthetic Protein Polymer  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,411A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/609,716  
; FILING DATE: 06-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/269,429  
; FILING DATE: 09-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/114,618  
; FILING DATE: 29-OCT-1987  
; PRIOR APPLICATION DATA:

APPLICANT: Ferrari, Franco A.  
FILING DATE: 04-NOV-1986  
TITLE OF INVENTION: Functional Recombinantly Prepared  
TITLE OF INVENTION: Synthetic Protein Polymer  
NUMBER OF SEQUENCES: 119  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,029A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/609,716  
FILING DATE: 06-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,429  
FILING DATE: 09-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-478-029A-65

Query Match 4.3%; Score 209; DB 4; Length 408;  
Best Local Similarity 23.1%; Pred. No. 1.5e-06;  
Matches 145; Conservative 14; Mismatches 200; Indels 270; Gaps 29;

QY 62 VGVQRRS-----RTCOLPTVQLHPSL-----PLPPRRHPEALLPRGQGRPQTS 107  
DB 4 VVLRDWNENPGVTQLNRLAAHPFASDPMGAPGPPGPPGPPGAGPPGP-PGPPGP 62  
QY 108 PETLPLXRTQSRGGRGLRGPASHLGRRETOEIRAARSRLRDPKPGMFGYGRVFFALP 167  
DB 63 P-----GPAGVGSFGA-----PGPPGPPG-----PGPPGPPG----- 82  
QY 168 LHRNRHRRSPRSSELSLSSRGEAISPTRAPRPFSSANGSPQTELPTELSTVHTPSQ 227  
DB 83 -----PPGPP-----GAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 123  
QY 228 ABPLSPETAQTEVAPTRPAPLRHHHPRAQASGTETPPSTHSLGEGFFRASPPRRPSQ 287  
DB 124 GPPGAP-----GPPGPPGP-----PGPPGPPGPPGPPGPPGPPGPPGPPGPP 159  
QY 288 GWASPOVAGRRDPFVSPPRGQQG-QGPWGTGTGPHGRLEPDQHPGAWLPLLSNGP 346  
DB 160 GPPGPPGA---PGP-PGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 215  
QY 347 -----HASSLWSLFAPSSIFPCSESEOLRACSQAPCEQPDPRALCOAFAFNSEFMG 401  
DB 216 PGPPGPPGPPGPPGPPGPPGPP-----PGPPGPPGPPGPPGPP----- 250  
QY 402 QLYOWEPFTEVQSRQCELNCRPGRFVYRHTKVDGTLQCGAPADICVAGRCLSPGC 461  
DB 251 -----PPG-----PPG-----PPGPPG---PAGVGSFGA 269  
QY 462 DGILSGRRPDGCGVCGDDSTCRVSGNLTDRGGPLGKQLWIPAGALRLQIAQLRPS 521  
DB 270 PGPPGPPGPPGPPGPPGPP-----GPPG-----GPPG----- 296  
QY 522 SNYLAIRGCGRSIINGNNAVDPPGYSRAGGTVFYRNRPPREGKESLSAEGPTTQPYD 581  
DB 297 -----PPGAPVGSFGAPGPP-----PPGPPGPPGAPGPPGPPGPP-- 333  
QY 582 VYMFQENBENGVFYQYVSSPP-----PILENPTPEPPVQLOPEILRVPPPLAPAPRAR 637  
DB 334 -----GPPGPPGPPGPPGPPGPPGPP-----GPPGPPGPP----- 365  
QY 638 TPGTLQROVRIPQMPAPPHPTPLGSPAA 666  
DB 366 -PGP-----PGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 386

RESULT 14  
US-08-478-029A-65  
; Sequence 65, Application US/08478029A  
; Patent No. 6184348  
; GENERAL INFORMATION:

```

; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21

Query Match      4.2%; Score 205.5; DB 3; Length 1078;
Best Local Similarity 22.5%; Pred. No. 7.8e+06;
Matches 206; Conservative 48; Mismatches 286; Indels 375; Gaps 52;

QY 40 GQPEGVWGFWOWASCSQCGVGVGVRRTCOLPTVOLHPSLPLPRP--PRHPEALL- 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 GPGSDGKPGP-----PQSQGESGRPGPGSPGPGQPGVGMG 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 97 ---PRGQ-----GPRPQTSPTLLPYRTQ-----SRGRGGPLRGPA 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 FPGFKGNDGAPKNGERGGPGGPGCPQPGPKNGEYQGQPPGTGPGGDGSDTGP-RGPPQ 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 SHLGREETQETIRAAKRSRLRDPITKPMFGYGVVPFALPLHNRHNP--RSPPRSLSLIS 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 GLOGLPCT-----GGPPGENEKPEPGPKGAGAPGAPGGKGADGAGPGERGPP-----GLAG 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 188 S----RGEAIPST-----PRAEPESANGSPQTE-LPPTELSHVHPSPQAEPLSPETA 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 APLRGAGGPPGEGGKGAGAPGCPP--GAAGTFCLOQNGPGERGLSGSPGKDGKGP-- 600
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

601	----	GP	CA	DV	PG	KD	GR	GT	PI	GP	PG	AG	PD	KD	GE	GG	----	AP	GL	PG	I	AG	PR	647																				
294	VA	-GRR	----	PD	PP	SV	----	PR	GR	QO	-Q	PP	WT	GT	GP	----	HG	PR	LE	P	330																							
648	GS	P	GE	RT	GP	PG	PA	FG	AP	QO	NG	EP	GG	K	ER	GA	PE	K	EG	GG	PG	VA	VP	PGG	SG	PA	GP	707																
331	DP	Q	----	HP	GA	WL	PL	LS	NG	PH	AS	LS	WF	AP	SS	PI	PR	CS	GE	SE	OL	RA	CS	QA	PC	PP	--	381																
708	GP	QV	K	GE	R	SG	P	G	----	GP	CA	A	G	----	FP	GA	R	L	P	GP	PG	SG	NG	----	PG	PP	GP	748																
382	----	EQ	PD	PR	AL	QA	AF	NS	OE	FM	GG	LY	QW	EP	TE	VQ	GS	OR	CE	LN	CR	PR	CF	RY	VR	HE	435																	
749	SG	SP	GD	PP	GA	NT	GA	PS	GV	SG	----	PK	DA	----	GO	PG	784																											
436	KV	OD	TL	CQ	GA	P	-DI	CV	AG	----	RCL	----	SP	CD	GI	L	S	GR	RP	D	CG	VC	GG	479																				
785	KG	SG	AO	GP	CA	GP	PL	GI	AG	IT	GA	R	LA	GP	MG	PM	PG	PR	GS	PG	QV	K	GE	SG	K	PA	NG	LS	- 843															
480	DD	ST	CR	LV	SG	N	L	T	DR	--	GP	LG	Y	O	K	IL	W	P	AG	AL	R	L	Q	I	A	LR	P	SS	N	Y	L	A	L	R	GP	G	R	SI	I	N	537			
844	----	ER	GP	PG	Q	----	LP	----	GL	AG	T	A	GE	PG	RD	866																												
538	GN	WA	VD	PP	GS	YR	AG	TV	FR	YR	NR	PR	EE	K	GE	SL	SA	EG	TT	QV	D	Y	M	I	F	Q	EN	VE	F	Y	QY	597												
867	GN	----	PG	S	----	D	L	GR	D	GS	PG	K	D	RG	EN	SG	FA	GA	----	PG	----	900																						
598	VI	SS	PP	PI	EN	TP	TE	PP	V	Q	L	PE	IL	VE	P	-PL	AP	-PR	AP	T	PT	GL	Q	R	V	I	P	OM	PA	P	654													
901	----	HP	GP	PG	V	GP	AG	K	SD	RG	ES	GP	AG	PA	GP	AG	PA	SG	RA	----	PG	940																						
655	PH	RT	PL	GS	P	----	AY	WK	RV	GH	SA	CS	AS	CK	GV	WR	IF	LC	IS	RE	SG	EL	DE	RS	CA	AG	710																	
941	Q	P	RG	D	K	GE	T	GE	R	GA	GI	K	----	GH	RG	FP	GN	P	----	AP	GS	974																						
711	PP	AS	PE	CH	GT	PC	PP	Y	WE	AG	EW	T	SC	RS	CG	PG	T	HR	L	Q	CR	QE	F	GG	G	SS	V	PP	ER	C	H	LP	770											
975	GP	AG	Q	GA	IG	SG	P	----	AG	PR	----	GP	V	GS	PP	GP	K	D	T	SG	1008																							
771	RP	NI	T	S	Q	CL	R	L	CH	WE	V	GS	PH	S	Q	VS	CR	G	R	Q	R	S	R	Q	Y	R	C	V	C	N	N	G	D	E	Y	S	----	Q	E	C	A	S	G	826
1009	H	F	----	PI	GP	----	G	P	----	G	P	----	G	N	R	G	E	R	G	E	S	G	P	H	PG	OS	G	1038																
827	PP	Q	P	S	R	E	A	C	D	M	G	P	C	841																														
1039	PP	GP	GA	P	----	G	P	C	1049																																			

AS 10 01 17 10 17 10

WED JUL 24 11:39:20 2002

Search completed: July 23, 2002, 21:05:25  
Job time: 2854 sec



OM of: US-10-041-770-2 to: Issued\_Patents\_NA.\* out\_format : pfs  
Date: Jul 23, 2002 9:42 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-framet\_p2n.model -Dev=xlh  
-Q/cgn2\_1/USP10.spool/US10041770/runat\_22072002\_153136\_2701/app\_query.fasta\_1.947  
-DB-Issued\_Patents\_NA -CFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bgsum62  
-TRANS-hunan40 cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10041770.ecgn1\_1\_59 -NCPG=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-10-041-770-2  
Query length: 877  
Database: Issued\_Patents\_NA.\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 54.460000

## score\_list:

Sequence	Strd Orig	ZScore	EScore Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-7	+	368.50	231.72	44377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1	+	368.50	231.72	44377
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-945-283-1	-	368.00	243.78	8438
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	-	359.50	192.73	4411529
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-1	+	355.00	230.01	15872
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-506-296B-3	+	347.00	234.52	3991
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-320-878-19	+	347.50	217.28	38506
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2	-	344.50	182.65	4403765
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-13	-	342.00	222.18	13987
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-7	-	342.00	213.91	44377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1	-	342.00	213.91	44377
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-32	+	341.50	223.42	11220
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-5	+	341.50	214.92	36778
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11	-	341.00	222.60	12001
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11	-	338.00	220.59	12001
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-428-517-1	+	337.50	209.89	50937
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-690-473-1	+	336.50	227.00	4257
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-821A-1	+	336.50	227.00	5.9e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-659-1	+	336.50	227.00	5.9e-05
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-384-490-30	+	335.50	223.75	2640
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-459-383-30	+	335.50	223.75	2640
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-537-002A-5	+	334.50	226.86	3600
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-863-010-5	+	334.50	226.86	3600
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-024-429-5	+	334.50	226.86	3600
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-3	-	331.00	214.98	13613
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-125-468-1	-	329.00	207.97	30001
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-474-933-1	-	329.00	207.97	30001
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-773-816-1	-	327.50	208.66	23673
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-335-409-1	-	327.50	201.02	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-102-1	-	327.50	201.02	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-567-969-1	-	327.50	201.02	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-480-1	-	327.50	201.02	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-486-1	-	327.50	201.02	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-472-1	-	327.50	201.02	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-36	+	325.00	219.64	4041
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-804-227C-1	+	323.50	201.65	43280
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1	+	323.50	219.98	3048
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-14418-1	+	322.50	219.98	0.0001
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-15327-1	+	322.50	219.98	0.0001
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-105-537-30	+	322.00	208.81	13842
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-977-434-9	+	320.50	220.04	2505

/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-458-819-9 + 320.50 220.04 0.0001 2505  
/cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:PCT-US91-07035-9 + 320.50 220.04 0.0001 25  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-643-734C-3 - 319.50 204.41 0.0011 20235  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-439-009A-3 - 319.50 204.41 0.0011 20235  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-804-227C-1 - 319.50 198.96 0.0021 43280

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-804-227C-7

## seq documentation block:

Sequence 7, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

alignment\_scores: Quality: 368.50 Length: 1067  
Ratio: 0.899 Gaps: 59  
Percent Similarity: 38.425 Percent Identity: 23.524  
alignment\_block:  
US-10-041-770-2 x US-08-804-227C-7

```

Align seg 1/1 to: US-08-804-227C-7 from: 1 to: 44377
53 TrpAlaserCysSerGlnProCysGlyValGlyValGlnArgSerAr 69
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29210 TGGCCCGCTGTGGTGGCCGACC...ACCGGCTGCTGCCCGCGTCCG 29256
59 gThrCysGlnLeuProThrValGlnLeuHisProSerLeuProLeuProp 86
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29257 GACGGCTGGT...CCTACG.....CCACGCCCGC 29282
86 roArgProProArg.HisProGlnAlaLeuLeuProArgGlyGlnGlyPr 102
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29283 CTCGGTCCCGGGGTTCCTCAGCGCTACTACGGTGTGGTCAACCTGG 29332
102 oArgProGlnThrSerProGluThrLeuProLeuTyrArgThrGlnSerA 119
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29333 CCGGCTCAGCGGGGAGCGGGTGTCTGTGCACGCCCGCGCGGGGC 29382
119 rg.....GlyArgGlyGly.....ProLeuArgGlyProAla 129
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29383 GTCGGCATGGCCCGGTGGATGCGGCCGCCACTCGGCGGGAGGTGCT 29432
130 SerHisLeuGlyArgGluGluThrGlnGluIleArgAlaAlaArgse 146
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29433 GGCCAC.....CGGAGCCCGGAAGTGGACGCCCTCGCGGCCATGG 29476
146 rArgLeuArgAspProIleLysProGlyMetPheGlyTyrGlyArgValP 163
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29477 GCATCACCAGCAGACCA.....LeuHisArgAsnArg 172
163 roPheAlaLeuPro.....LeuHisArgAsnArg 172
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29494 CTCGCCTCTCCCGCACCTCGACTTCGCGACGCGCTTCACCGAGCGGA 29543
173 ArgHisProArgSerProProArgSerGluLeuSerLeuIleSerSerAr 189
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29544 CGGCACGTCGCCGGGAGTGTCTCTGAACCTCGCTCACCAAGAGGTGC 29593
189 gGlyGluGluAlaIleProSerProThrProArgAlaGluProPheSerA 206
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29594 TGGAGC.....CTCCCTCGGGCTGCTCGTCCGGGGCGCGTTCCTG 29637
206 laAsnGlySerProGlnThrGluLeuProProThrGluLeuSerValHis 222
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29638 GAGCTGGGCAAGACCCAGCTCCGGGACCCCGAGCGATCCCGG...CGA 29684
223 ThrProSerProGlnAlaGluProLeuSerProGluThrAlaGlnThrGl 239
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29685 ACACCCCGGGTGGCGTACGGGCGGTTCACCTCACGAGCGCGGACCCG 29734
239 uValAlaProArgThrArgProAla..... 247
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29735 A.....CGCAGCTCGCGGCTGTGGGAACTGATGGACCTGTC 29775
248 .....ProLeuArgHisProArgAlaGlnAla 257
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29776 GCGCGCGGCTGTGCACCGCTGCCCGTGTCAACCCACGACGTGGCGCG 29825
258 SerGlyThrGluPro.....ProSerProThrHisSerLeu.. 269
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29826 GGCGCGGAGCGCTCGCACCATCAGCCAGCGCGGCACACCGGAAGC 29875
270 .....GlyGluGly 273
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29876 TGTCTGTGACATGCGCGCGCTGGCACCCGTACGGCAGCGTCTGGTC 29925
273 lyPhePheArgAlaSerPro...GlnProArgArgProSerSerGlnGly 288
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29926 ACCGGTGGACCGCGGCTCTGGCAGCGCATCGCCG..... 29963
289 TrpAlaserProGlnValAlaGlyArgPro.....AspProPh 302
||||| :||| :||| :||| :||| :||| :||| :||| :|||
29964 .....CCACTGGGAGCGGACCGCGTCCGCGCGTGTGTATCGCG 30007

```

[illegible]

[illegible]



```

; TELEPHONE: 317-276-3865
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1
;
alignment_scores:
; Quality: 368.50 Length: 1067
; Ratio: 0.899 Gaps: 59
; Percent Similarity: 38.425 Percent Identity: 23.524
;
alignment_block:
US-10-041-770-2 x US-08-804-198-1 ..
;
Align seg 1/1 to: US-08-804-198-1 from: 1 to: 44377
;
53 TrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgArgSerAr 69
||||| ||||| ||||| ||||| |||||
29210 TGGGCGCGCTGGCTGGCCGACC...ACCGCTGCTCGCCCGCTCCG 29256
;
69 gThrCysGlnLeuProThrValGlnLeuHisProSerLeuProLeuProp 86
||||| ||||| ||||| ||||| |||||
29257 GACGCTGGT...CCTACG.....CCAGGCGC 29282
;
86 roArgProProArg.HisProGluAlaLeuLeuProArgGlyGlnGlyPr 102
||||| ||||| ||||| ||||| |||||
29283 CTCGGTCGCCGGGTGTTCTTCAGCGCTACTACGGTGTGTCACCTGG 29332
;
102 oArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSerA 119
||||| ||||| ||||| ||||| |||||
29333 CGGGCTCAGCGCGGGGAGCGGGTGCTCGTGACCGCGCCCGGGGGC 29382
;
119 rg.....GlyArgGlyGly.....ProLeuArgGlyProAla 129
||||| ||||| ||||| ||||| |||||
29383 GTCGCGATGGCGCGGTGCAGATCGCGCGCACCTCGCGCGGAGGTGCT 29432
;
130 SerHisLeuGlyArgGluGluThrGlnGluLeuArgAlaAlaArgSer 146
||||| ||||| ||||| ||||| |||||
29433 GGCCAC.....CGCAGCCCGGCAAGTGGGACGCCCTGGCGCCATGG 29476
;
146 rArgLeuArgaspProIleLysProGlyMetPheGlyTyArgValp 163
||||| ||||| ||||| ||||| |||||
29477 GCATCAGCAGCACCA.....C 29493
;
163 roPheAlaLeuPro.....LeuHisArgAsnArg 172
||||| ||||| ||||| ||||| |||||
29494 CTCGCGCTCTCCGCGACCTCGACTTCGCGACCGCTTCACCGGAGCGGA 29543
;
173 ArgHisProArgSerProProArgSerGlnLeuSerLeuIleSerSerAr 189
||||| ||||| ||||| ||||| |||||
29544 CGCAGCTCCCGCGGAGCTGCTCTCACTCACTCGTCACCAAGGATTCG 29593
;
189 ggGlyGluAlaIleProSerProThrProArgAlaGluProPheSerA 206

```

[illegible]

30392 .....CCGGCGCGGTCCCTCGGC 30411  
414 ySerGlnArgCysGluLeuAsn.....CysArgProArgGlyPheArgp 429  
:|||||:|||||:|  
30412 CTCCTGCTGTCACCTCGCTGGGTCTGTGTCGACACACCGCGGAT 30461  
429 heTyrValArgHisThrGluLysValGlnAspGlyThrLeu.....442  
:|||||:|||||:|||||:|||||:|  
30462 GGCACACGACTGGACAGAGAGCCTCCAGCGGCCTTCGCCCGCAGCG 30511  
443 .....CysGlnPro...GlyAlaProAspIleCysValAlaG1 454  
:|||||:|||||:|||||:|  
30512 GCTTCCCGCCCTGTCCGCCACGCTGGGCGCG.....30544  
454 yArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgp 471  
:|||||:|||||:|||||:|  
30545 .CGCTGTCGACGCCCTGTC...GGTTCGACGAGCGCTGCAGGTCCC 30590  
471 roAspGly.....CysGlyValCysGlyGlyAspAspSer 482  
:|||||:|||||:|||||:|  
30591 CATCGGTTCGACCCCGCGCGCTGCGCGCCACCGGAAGCTCCCGGCC 30640  
483 ThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuG1 499  
:|||||:|||||:|||||:|  
30641 TGCTGTCGACCTCGTGGGTCCG...CCCGGCGACCGGTCCG.....30682  
499 yTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleA 516  
30683 .....C 30683  
516 laGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGly 532  
:|||||:|||||:|  
30684 GGCCCGCGCTCGGCCCTTCCTCGGCTCGGACGCGCGGACCGTGGCG 30733  
533 ArgSerIleLeuAsnGlyAsnTrpAlaValAspProProGlySerTyrAr 549  
:|||||:|||||:|||||:|  
30734 AGCGCTCGCGGAGCGGT...TGGCGGAGCTCTCCCGGAGGAAACCCAC 30780  
549 gAlaGlyGlyThrValPheArgTyrAsnArg.....559  
30781 GACCGGTCTCGGCTGGTTCGCGAACACGTGGCGCGCTACTGGGCCA 30830  
560 .....ProProArgGluGluGlyLysGlyGluSerLeuSerAlaGlu 573  
:|||||:|||||:|  
30831 CGGCTCGCGCGGAGTTCGCGCGACCGCGGTTCGCGAGTGGGT 30880  
574 GlyProThrThrGlnProValAspValTyrMetIlePheGlnGluAs 590  
:|||||:|||||:|  
30881 TCGACTCGTACGCGCGGTGGAACTGGCCACCGGATGGCG...CGGTC 30927  
590 nProGlyValPheTyrGlnTyrValIleSerSerProProIleLeuG 607  
:|||||:|||||:|  
30928 ACCGGG.....TCAGGCTCCCGCGACCCCTGGTCTT 30959  
607 luAsnProThrProGluProProValProGlnLeuGlnProGluIle... 622  
:|||||:|||||:|  
30960 CGACCAACCCACCCCGCGCTGCTCTCGACCTCGAGCGGCTGTGG 31009  
622 ..... 622  
31010 CCCCGCACAGCGGTACACACACACCGGTGCTGTCGGAACGGACCGC 31059  
623 ....LeuArgValGluProProLeuAlaProAlaProArgProAla... 636  
:|||||:|||||:|  
31060 ATCAGAGAGGCGCTGGCGGCTTCACCCCGAGACCTCGCGAGGTGCG 31109  
637 .....ArgThrProGlyThrLeuGlnArgGlnValArgIleProG1 650  
:|||||:|||||:|  
31110 CCCCGCCCGACGACCGGCGGA.....GG 31135  
650 nMetProAlaProProHisProArgThrProLeuGlySerProAla... 665  
:|||||:|||||:|  
31136 TCGCCCTGCGCTGGACCGCTGCGGACCGCTGGCGCGCTGCACGAC 31185

665 ..... 665  
31186 GCGCGCCCGCGCGGACGACGACATCACGACGTGCTGACGAGCGCGA 31235  
666 .....AlaTyrTrpLysArgValGlyHisSerAlaCy 676  
:|||||:|||||:|  
31236 CGACGACGAGATCTTCGGCTTATCGAGCGCGGTACGCGACGTGCTGAC 31285  
676 sser.....AlaSerCysGlyLysGlyVal.TrpArg 686  
:|||||:|||||:|  
31286 GCGCGCCCGGAGCCCGCGCTCATCGAAAGAACACCACTGGCGA 31335  
686 ..... 686  
31336 AGAAGAGAGCTGCGCGCTACCTCAAGCGGTGACGGGTGAGCTGCAC 31385  
687 ....ProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAspG 702  
:|||||:|||||:|  
31386 CGGCGCACCGACGACGTGCTGCTGGACCGGCGGCGCCACGAGCCGAT 31435  
702 luArgSerCysAlaAlaGlyAlaArgProProAlaSerProGlu..... 716  
:|||||:|||||:|  
31436 CGCGATCTCGGCGCGCTGCGGACTCCCGGCGGCTCGAGAGTCCGG 31485  
717 ...ProCysHisGlyThrProCys..... 723  
:|||||:|||||:|  
31486 AGGACTGT...GGAGCTGTGACGCGGTGCGGACGCGTGGCGCG 31532  
724 .ProProTyrTrpGluAlaGlyLysThrSerCysSerArgSerCys. 739  
:|||||:|||||:|  
31533 GCGCCCGCGACCGCGGTGGACGTGAGGAAAGTACTCGCCGACCC 31582  
740 .....GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGln 752  
:|||||:|||||:|  
31583 CGACACGCGCGCACCTCGTACTGCGCGGAGGCGGCTTCGTGAGG... 31629  
753 GluPheGlyGlyGlySerSerValProProGluArgCysGlyHisLe 769  
:|||||:|||||:|  
31630 .....GGCGGACCGGTTCGACCCCGCTCTTCGCGATCTCGGCC 31670  
769 uproArgProasn.....IleThrGlnSerCysGlnLeuArgLeuC 783  
:|||||:|||||:|  
31671 AACGAGGCGCTCACCATGACCCCGAGCGGCTGCTGGAGACCTC 31720  
783 ysGly...HisTrpGlu.....ValGlySerProTrpSerGln 794  
:|||||:|||||:|  
31721 CTGGAGGCGCTGGAGCGCGGTCTGGACCCCGCTCTCCGCGGCA 31770  
795 CysSerValArgCys.....GlyArgGlyGlnArgSerArgGlnValar 809  
:|||||:|||||:|  
31771 GCGGACCGCGCTGTTCGCGGCGGTGGAGAGCGGTACACGAGAGGC 31820  
809 gCysValGlyAsn.....AsnGlyAspGluValSerG 820  
:|||||:|||||:|  
31821 GTCGAAGGCTCGAAGCGCATCTGAGGCGCCCACTCTGCGCGGCTCGT 31870  
820 luGlnGluCysAlaSerGlyProProGlnPro.....ProSerArg 833  
:|||||:|||||:|  
31871 CAGCTTACCGCGCGCGCTGCTGCTACGCTGCGGCTGAGGCGCGG 31920  
834 GluAlaCysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTr 850  
:|||||:|||||:|  
31921 CCTGACGATCGACAGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 31970  
850 p..... 850  
31971 GCGGTGACGTCACCTGCGCGGCGGAGTGCAGACCTCGCACTGCGCGCGG 32020  
851 .....SerSerLysValSerProGluProPro 859  
:|||||:|||||:|  
32021 GCGCAGGCTCATCGCGGACTTCGCGCTTTCACCCAGTTCTCCC 32064

seq\_name: /sgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-07-945-283-1

```
seq_documentation_block:
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EPO and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL USA
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P.
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1267, "t")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
; US-07-945-283-1
```

alignment\_scores:  
Quality: 368.00 Length: 930  
Ratio: 1.005 Gaps: 46  
Percent Similarity: 39.355 Percent Identity: 24.839

alignment\_block:  
US-10-041-770-2 x US-07-945-283-1/rv ..

Align seg 1/1 to reverse of: US-07-945-283-1 from: 1 to: 8438

```
30 GlyHisSerLeuGlnThrPro.....ThrGluGluGlyGlnG 42
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6437 GGGCTCTCCATCAGGGCTCCGAGCAGCAGGTCCACCTCGCTCGGCGCCGG 6388

42 yProGluGlyValTyrGlyProTyrValGlnTyrAlaSerCysSerGlnP 59
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6387 TCCGGGGCGGCTCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6338

59 roCysGlyValGlyValGlnAlaArgSerArgThrCysGlnLeuProThr 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6337 GGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6288

76 ValGlnLeuHisProSerLeuProLeuProProArgProProArgHisPr 92
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6287 CCTCCCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6238

92 oGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThrSerProG 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6237 GCGCCCTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6192

109 luThrLeuProLeuTyrArgThrGlnSerArgGlyArgGlyGlyProLeu 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6191 .....CGGTCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 6159

126 ArgGlyProAlaSerHisLeuGlyArgGluThrGlnGluIleArgAl 142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6158 ...GGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6121

142 alaAlaArgSerArgLeuArgAspProIleLysProGlyMetPheGlyT 159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6120 GCAGCGCGCG.....CACGCGCG.....CACGCGCG 6104

159 yrGlyArgValProPheAlaLeuProLeuHisArgAsnArgHisPro 175
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6103 GGGCGCGGCTTCCCGAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 6054

176 .....ArgSerProArgSerGlu.....LeuSerLeuIleSe 187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6053 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6004

187 rSerArgGlyGluAlaIleProSerProThrProArgAlaGluProp 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6003 CAGCAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5954

204 heSerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSer 220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5953 TCCTCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5904

221 ValHisThrProSerProGlnAla.....GluProLeuSerPr 233
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5903 GCTCCAGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5854

233 oGluThrAlaGlnThrGluValAla...ProArgThrArgProAlaProL 249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5853 GAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5804

249 euArgHisProArgAlaGlnAlaSerGlyThrGluPro..... 262
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5803 GCAGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5754

263 ProSerProThrHisSerLeuGlyGluGlyPhePheArgAlaSerPr 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5753 GCGCGGCGCTTACGACGCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 5704

279 oGlnProArgProProSerSerGlnGly.....TrpAlaSerProGlnV 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5703 CCGGCGCGCTCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5654

294 aAlaGlyArgArgProAspProPheProSerValProArg..... 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5653 CGCGCGCGCGCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5604
```

```
307 .....LysGlyGluSerLeuSerAlaGluGlyPro...ThrThrG 578
5603 CCTGCTCTCTCCGAGGAGAGAGAGAGAGAGTCCGGCCGGCGGC 5554
308 .....GlyArgGlyGln...GlnGlyGlnGlyProTyrP 318
5553 CCACCTCCCGCCGCGCGCGCCGACCGCTCCCGCCGCGCGCGCGCGC 5504
318 lyThrGlyGlyThrProHisGlyProArgLeuGluProAspPProGlnHis 334
5503 CCGCGCGCCCTCGGCTCTCCGCTCCGCGCACCTCTCTCTCCGCGCGG 5454
335 ProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisSerSerLe 351
5453 CCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 5404
351 u.....TyrSerLeuPheAlaProSerS 359
5403 CGCTCCACCAACACCACTCTCGCTCATGCGCGAGC...GGCCCGCGCC 5357
359 erProIleProArgCysSerGlyGluSerGlnGlnLeuArgAlaCysSer 375
5356 GACCGAGCGCGCGCTGCTCACCC.....CGCTCGGGGAGC 5322
376 GlnAlaProCysProGluGlnProAspProArgAlaLeuGlnCysAl 392
5321 CCTGGCGCGCTCCGACCTCCGCGCGAGCGCGCGCTCCGCTACGGCGG 5272
392 aAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrGlnTrpGluProP 409
5271 G.....CCGG 5267
409 heThrGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProArg 425
5266 GGACTCCCGCGAGGGCTCTGGGACGAGGAGGAGCTCCGCGAGCGCGG 5217
426 GlyPheArgPheTyrValArgHisThrGluLysValGlnAspGlyThrLe 442
5216 .....CCCGCTACCGC 5206
442 ucysGlnProGlyAlaProAspIleCys..... 451
5205 GCGCGCGCGCGCGCG.....TGCCGCTCTCATCCCGGAGATGGG 5165
452 ..ValAlaGlyArgCysLeuSerProGlyCys...AspGlyIleLeuGly 466
5164 GGACTCGAGAGAGCAGCAGCAGCGCTCGTCCGCTCATCTACAGCGGG 5115
467 SerGlyArgArgPro...AspGlyCysGlyValCysGlyGlyAspAspSe 482
5114 CCGCGCGCGAGGCGCATGCTCTGGCTGCAGAACCCCGGA..... 5076
482 rThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuG 499
5075 ....TGCAGGCCCCCGACCGCTTCAACCCAGTCTGCCAGC..... 5037
499 lyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIle 515
5036 .....GCCGGTCCACGCG.....CCC 5019
516 AlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyG 532
5018 ACGGCACCGCTCTCTCATCACC..... 4995
532 yArgSerIleLeuAsnGlyAsnTrpAlaVal..AspProProGlySerTyr 548
4994 .....GCAGCGTACCGCGCGCGCTGCCGCA 4970
549 ArgAlaGlyGlyThrValPheArgTyrAsnArgProProArgGluGluG 565
4969 CATCGGGGA.....CGCATGCGCGCGCGCGCGCTCTGGG 4932
```

```
565 Y.....LysGlyGluSerLeuSerAlaGluGlyPro...ThrThrG 578
4931 CCTGCGCGACGCGGTGAGCGCGCTGCGCATGAGCCGCGTACGATCGC 4882
578 lnProValAspValTyrMetIlePheGlnGluGluAsnProGlyValPhe 594
4881 ACCAGAGACCTTCATCT..... 4862
595 TyrGlnTyrValIleSerSerProProIleLeuGluAsnProThr... 610
4861 .....GCAGAGCTCCGCGCGCGCTACG...CCGACATG 4831
611 .....ProGluProProValProGlnLeuGlnProGluI 622
4830 GCCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4782
622 leuArgValGluProProLeuAlaProAlaProArgProAlaArgThr 638
4781 .....AGGCCCTCTCGCGCGCGCTCGCGCGCGCTTCGCGCGCGCGC 4737
639 ProGlyThrLeuGlnArgGlnValArgIleProGln..... 650
4736 CCGCGCGCTGCGCGCGAGCTGGCGAGCGCTGCTGCTGCGCTGCGCG 4687
651 .....MetProAlaProp 655
4686 GCGCTCTCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4637
655 rHisProArgThrProLeuGlySerProAlaAlaTyrTrpLysArgVal 671
4636 CCGCGCGCGCGCGCGCTGCGCGCGCGCTGCGCTGCGCTGCGCTG 4596
672 GlyHisSerAlaCysSerAlaSerCysGlyLys.....GlyValTrpAr 686
4595 TGACCGCGCGCTGCTGCGCGCTGCGCGAGCGATCCCGCGCGCGCGCG 4546
686 gProfilePheLeuCysIleSerArgGluSerGlyGluGluLeuAspGlu 703
4545 GCG..... 4542
703 rgSerCysAlaAlaGlyAlaArgPro.....ProAlaSerPro 715
4541 .....AGCGCAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 4503
716 GluProCysHisGlyThrProCysProProTyrTrpGluAlaGlyGlu 732
4502 GCACCTGCGC.....CGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4471
732 pThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuG 749
4470 GAGCGCGCGCGCGCGCGCGCGCGCTGAGCTACGCGCGCGCGCTCT 4421
749 lnCysArgGlnGluPheGlyGlyGlySerSerValProProGluArg 765
4420 CGCCCGCGCAACGTGCGCGCGCGCGCTGCGCGAGCGCGCGCGCGCG 4371
766 CysGly...HisLeuProArgProAsnIleThrGlnSerCysGlnLeuAr 781
4370 CCGCGCGCGCGAGCGCGCG.....GCCTGC..... 4341
781 gLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSerVal 798
4340 .....CCCGCTCTGCGCGCGAGCAGC 4319
798 rgCysGlyArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsn 814
4318 CGGCTCTGCTCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4269
815 GlyAspGluValSerGluGlnGluCysAlaSerGlyProProGlnProPr 831
4268 GCTCCCGG.....GCTCCGCGCGCGCTCCAGCGCGCGC 4237
831 oSerArgGluAlaCysAspMetGlyProCysThrThrAlaTrpPheHis 848
```

4236 AGCACAAAGTCGGCTCCAGCACCAGTCCAGTCCGGCA..... 4197  
848 erAspTrpSerSerIysValSerProGluProProAla 860  
4196 .....CAAAGTCGGCTCCGGCTCCGCTCCGGCT 4167  
seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-103-840A-1

seq\_documentation\_block:  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103.840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

alignment\_scores:  
Quality: 359.50 Length: 938  
Ratio: 0.999 Gaps: 50  
Percent Similarity: 38.380 Percent Identity: 25.160

alignment\_block:

US-10-041-770-2 x US-09-103-840A-1/rev ..  
Align seq 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529  
74 ProThrValGlnLeuHisProSerLeuProProArpProAr 90  
9394104CCGGCGCGCGCTGCGCGCGCGCGCTTGCGCGCTTGGCGCGTC 3934055  
90 gHisProGluAlaLeuProArpGlyGlnGlyProArpProGlnThrs 107  
9394054AAAGCGGAGAGCCAGCGCGAGACCGCTGCGCCCATCGCGCGCTCAC 3934005  
107 exProGlu.....ThrLeuProLeuTyr.Argth 116  
9394004CGCGCGCGCGCTTCCCGCGCGCGCTTACCGCGCGCGCGCG 3933955  
116 rGlnSerArgGlyGlyProLeuArgGlyProAlaSerHisLeuG 133  
9393954AAGCGCGCGCACACAGCGCGCGATCGC..... 3933924  
133 lyArgGluGluThrGlnGluLeuArgAlaArgSerArgLeuArg 149  
9393923.....CGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933888  
150 AspProLeuSerProGlyMetPheGlyTyrGlyArgValProPheAlaLe 166  
9393887.....CCAGGG.....TT 3933880  
166 uproLeuHisArgAsnArg.....ArgHis..ProArgSerProAr 180  
9393879CCCGACACCGCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCG 3933830  
180 gserGluLeuSerLeuHisSerSerArgGlyGluGluAlaLeProSerP 197  
9393829GCGCGCGCTTGCGCTTGATC.....GAGGTGGTGTGCGCGCG 3933792

197 roThrProArgAlaGluProPheSerAlaAsnGlySerProGlnThrGlu 213  
3933791CCAGCGCGCGCTTGGCGCGCTGG.....CCGCGCGCGCACCC 3933754  
214 LeuProProThrGlu...LeuSerValHisThrProSerProGlnAlaG 229  
3933753TGGCGCGCGCTTGAATTTGTTCCCGCGCGCGCGCGCGCGCG 3933704  
229 uproLeuSerPro...GluThrAlaGlnThrGluValAlaProArp 245  
3933703CCCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCTG 3933654  
245 rgProAlaPro...LeuArgHisHisProArgAlaGlnAlaSerGlyThr 260  
3933653GGCGCGCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCG 3933604  
261 GluProProThrProThrHisSerLeuGlyGluGlyGlyPhePheArg 277  
3933603TCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933584  
277 aserProGlnProArgArgProSerSerGlnGlyTyrAlaSerProGlnV 294  
3933583GTGCGCGAGACCTGCGCGCGCGGTTCGCGCGCGCGCGCGCG 3933534  
294 alaAlaGlyArgProAspPro...PheProSerValProArgGlyArg 309  
3933533TGC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933484  
310 GlyGlnGlnGlyGlnGlyProTyrGlyThrGlyGlyThrProHisGlyPr 326  
3933483CGCGCGCGCGCGCTTGGCGCGCTTG.....CGCGCGTCAAAGCGGAGAGCC 3933437  
326 oArgLeuGluPro.....AspProGlnHisProG 336  
3933436CAGCGCGAGACCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 3933387  
336 lValaTyrLeuProLeuSerAsnGlyProHisAlaSerSerLeuTyr 352  
3933386CGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCG 3933349  
353 SerLeuPheAlaProSerSerProLeuProArgCysSerGlyGluSerGl 369  
3933348.....CGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933308  
369 uGlnLeuArgAlaCysSerGlnAlaProCysProProGluGln..... 383  
3933307GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933258  
384 .....ProAspPro.ArgAlaLeuGlnCysAlaAlaPh 394  
3933257GCGTGGGTGTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933208  
394 easnSerGlnGluPheMetGlyGlnLeuTyrGlnTyrGluProPheThrG 411  
3933207C.....CGCTGACCGCGCGCGCGCGCGCGCGCGCGCG 3933179  
411 luValGlnGlySerGlnArgCysGluLeuAsnCysArgPro..... 424  
3933178GTTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933138  
425 .....ArgGlyPheArg..... 428  
3933137CTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933088  
429 .....PheTyrValArgHis..ThrGluLysValGlnAsp 439  
3933087GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933038  
440 GlyThrLeuCysGlnProGlyAlaProAspPheCys.....ValAlaG 454  
3933037GCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3932988

```
454 yArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgp 471
      :::::
3932987TGC CGCGGTACGATGGCGCCACCGCGCCCGCGCGGTGCGCGGTCCG 3932938
      :::::
471 roaspGlyCysGlyValCysGlyGlyAspasp..... 481
      :::::
3932937CCG GTCCGCGGTGCGCGCGCGCGCGCGCGATGCCG GTGGGTATCCGC 3932888
      :::::
482 .....Se 482
3932887GCGGCGCGCGCGCGCGCGCGCGCGCGCTTGTCCGCGGTGCGCGCAG 3932838
      :::::
482 rThrCysArgLeuVal..... 487
3932837CGCCACACAGACCGGTGTGTCGCCCGCGCACCGCTGCTACGCCCTGCCCG 3932788
      :::::
488 .....SerGlyAsnLeuThrAsp 493
3932787CCGCGCGCGCGCGCGCGCGCGCGCGCTTCCCGCGCGGTGCGCGCGAT 3932738
      :::::
494 ArgGlyGlyProLeuGlyTyrGlnLysIle..... 503
3932737GCCG GTGGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3932688
      :::::
504 .....LeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaG 517
3932687CGCGCGCGGTGCGCGCTTGC CGCGCGCGAGCGGT..... 3932654
517 lnLeuArgProSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArg 533
      :::::
3932653.....GTTGCGCGCAACCGCGGTGTGCGCGCGCGCGCGCGCGCGCG 3932612
      :::::
534 SerIleIleAsnGlyAsnTrpAlaValAspProGlySerTyrArgAl 550
      :::::
3932611GCCG GTCCACCGGANA.....CCCGGTGCC.....GC 3932583
550 aGlyGlyThrValPheArgTyrAsnArgProProArgGluGluGlyLysG 567
      :::::
3932582CGGTGCGCACCGGTAGCGTCCGTCGCGCGCGCTGCCCGCGCGCGCG 3932533
      :::::
567 lyGluSerLeuSer.....AlaGluGly.Pr 575
      :::::
3932532CGCGCGCGCTTGC CGCGCGCGGTGCCCGCGATACCGCGCTTGC CGCGCGCG 3932483
      :::::
575 oThrThrGlnProValAspValTyrMetIlePheGlnGluGluAsnProG 592
      :::::
3932482GACAAACCGCGGT..... 3932467
592 lyValPheTyrGlnTyrValIleSerSerProProIleLeuGluAsn 608
      :::::
3932466.....CCTCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3932449
609 ProThrProGluProProValProGlnLeuGlnPro..... 620
      :::::
3932448CCG...CGCGCGCGCGCGCGCGCGCGGTACCGCGCAGTCCCGCGCGCGCG 3932402
621 .....GluIleLeuArgValGluProProLeuA 630
3932401GCCGTGCGCGCAATCCGCTGGCATTATCAGCACCGGAGCACCGCATGC 3932352
      :::::
630 laProAlaProArgProAlaArgThrProGlyThrLeuGlnArgGlnVal 646
      :::::
3932351CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 3932302
      :::::
647 ArgIleProGlnMetProAlaProHisProArgThrProLeuGlyse 663
      :::::
3932301ATGCGCGCGCAATGGGTGTCGCCCTGGCGCGCGCGCGCGCGCGCGCGCG 3932252
      :::::
663 rProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerAlaSerC 680
      :::::
3932251GCCAGCGAACCGGTACCGCGGT..... 3932227
680 ysGlyLysGlyValTrpArgProIlePheLeuCysIleSerArgGluSer 696
```

seq\_name: /cogn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-105-537-1

```
seq_documentation_block:
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1
```

alignment\_scores:  
Quality: 355.00 Length: 1066  
Ratio: 0.822 Gaps: 61

[illegible]

```
504 LeuTrpIleProAla.....GlyAl 510
      |||:::
5836 CGCGTTCGACCGCGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 5885
510 aLeuArgLeuGlnIleAlaGlnLeuArgProSer.....SerAsnTyrLeuA 526
      |||:::
5886 CGCGCGCGGCGGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 5935
526 laLeuArgGly...ProGlyGly.....ArgSerIleIleAsnGlyAsn 539
      |||:::
5936 CCTGTGCTGCGCGGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 5985
540 TrpAlaValAspProProGlySerTyrArgAlaGlyGlyThrValPheAr 556
      |||:::
5986 CGCGCTT.....CCTGGAGTCTCCCTCGCGGCGGCGA..... 6018
556 gTyrAsnArgProProArgGluGluGluGlyGlyGluSerLeuSerAla 572
      |||:::
6019 .....CCAGTTCGGGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 6055
573 GluGlyProThrThrGlnPro.....ValAspValTyr 583
      |||:::
6056 GAAGCGCGGCGTGTGTGTGAGCACCAGGTGCGAGTGG 6105
583 rMetIlePheGlnGluGluAsnProGlyValPheTyrGlnTyrValIleS 600
      |||:::
6106 CGTACGCGGCGGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 6134
600 er.SerProProIleLeuLeuGluAsnProThrProGluProProValPr 616
      |||:::
6135 GCGCGACCTTGGTGTGTGTGAGCACCAGGTGCGAGTGG 6184
516 oGlnLeuGlnProGluIleLeuArgValGluProProLeuAlaProAla. 632
      |||:::
6185 GACGCGCGCGGCGGT.....GGACCGCGCATGCTCTCGGCGGT 6222
633 .....ProArgProAlaArgThrPro..... 639
6223 ACTCGCGGAGGCGGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 6272
640 GlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProProH 656
      |||:::
6273 GGGACGGTTCAGCGGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 6316
656 spro..... 657
6317 CCCCTGACGCGGCTGTACGACGCGGCGTGTGTGAGCACCAGGTGCGAGTGG 6366
658 .....Arg 658
6367 TCGGCGCTTCCAGGCGCTGACGGGCGTGTGTGAGCACCAGGTGCGAGTGG 6416
659 ThrPro...LeuGlySerProAlaAlaTyrTrpLysArgValGlyHisE 674
      |||:::
6417 TCGCGGAGATCCGCGTGTGTGAGCACCAGGTGCGAGTGG 6466
674 rAlaCysSerAlaSerCysGly.....LysGlyValTrpA 686
      |||:::
6467 GCGGTACACCGGCGGCTGTGTGAGCACCAGGTGCGAGTGG 6516
686 rg.....ProIlePheLeuCysIleSerArgGluSer 696
      |||:::
6517 GGGCAACTCAGCTCCGCGTGTGTGAGCACCAGGTGCGAGTGG 6566
697 GlyGluGluLeuAspGluArgSer..... 704
      |||:::
6567 GGTTCGCGTGTGTGTGTGAGCACCAGGTGCGAGTGG 6616
705 .....CysAlaAlaGlyAlaArgProProAlaSerProGluProC 718
      |||:::
6617 GCCACGCGGCTTCGCGTGTGTGAGCACCAGGTGCGAGTGG 6666

718 ysHisGlyThrProCysProProTyr..... 726
      |||:::
6667 GCTCGCGCTACCGACACACCGGTGTGTGAGCACCAGGTGCGAGTGG 6716
727 .....TrpGluAlaGlyGluTrp..... 732
      |||:::
6717 TCACCTCGCGGCGTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 6766
733 .....ThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgG 747
      |||:::
6767 GTGACGACGCGCGCTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 6816
747 InLeuGlnCysArgGlnGluPheGlyGlyGlySerValProPro 763
      |||:::
6817 ACCCGACGCGGCGGCGTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 6866
764 GluArgCysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLe 780
      |||:::
6867 GCTCG..... 6872
780 uArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysServ 797
      |||:::
6873 .....GCGACCTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 6897
797 alArgCysGlyArgGlyGlnArgSerArgGlnValArgCysValGlyAsn 813
      |||:::
6898 GACGCTGCTCTCCGCGGCGGCGTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 6932
814 AsnGlyAspGluValSerGluGlnCysAlaSerGlyProGlnPr 830
      |||:::
6933 ACGGGACGCGGCGGCGGCGTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 6982
830 oProSerArgGluAlaCysAspMetGlyProCysThrAlaTrpPheH 847
      |||:::
6983 ATCTGTCAGGAGTCTCTGCGGCGGCGTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 7022
847 isSerAspTrpSerSerLysValSer..... 855
      |||:::
7023 .....ProGluProAlaIleSer 862
7071 AGGAGTACCTCTCTGCGGCGGCGTGTGTGAGCACCAGGTGTGTGAGCACCAGGTGCGAGTGG 7103

seq_name: /cgn2_5/ptodata/2/ina/6B_COMB.seq:US-08-506-296B-3

seq_documentation_block:
; Sequence 3, Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crossin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,296B
; FILING DATE: 24-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
```



```

NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3991 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 59..3859
US-08-506-296B-3

```

alignment\_scores: Quality: 347.00 Length: 1103  
Ratio: 0.801 Gaps: 64  
Percent Similarity: 39.257 Percent Identity: 24.297

alignment\_block:  
US-10-041-770-2 x US-08-506-296B-3 ..  
Align seg 1/1 to: US-08-506-296B-3 from: 1 to: 3991

```

27 ValLeuSerGlyHisSerLeuGlnThrProThrGluGluGlyGlnGlyPr 43
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27 GTGAGACAGCGCGCGAGCTGCCGCTCCG..... 56
43 oGluGlyValTrpGlyProTrpVal...GlnTrpAlaSerCysSerGlnP 59
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 .....CCATGGCTCTCCCATGGTCGGCTCTCTCTGCT 90
59 roCysGlyValGlyValGlnArg..... 66
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
91 CCTGTGTGGGGGCGCGAGCGCCATCACCATTCCCGCGAGTATG 140
67 ...ArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro..... 80
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 GTGGCAGCATTCCTGACGCCCCCGAGCTGAGGAGGAACCCCGAA 190
81 .....SerLeuProLeu.....Prop 86
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 CAACTCGTGTCTTCCCGAGTGATGACATGCTCTCAATGCGTGGCCAC 240
86 roArgProProArg.....HisPro.....GluAlaLeuLe 96
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 CGGAACCCCGCTCCAGTACCATGAGCGGTGAGGATCAGCCCTTCG 290
96 uProArgGlyGln.....GlyProArg..... 103
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
291 TCCCGAGGACACGGGGGTCTCGTGTGTCTCCCGGATCGGGACTTG 340
104 .....ProGlnThrSerProGlnThrLeuProLeuTyr 114
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 GTCATCAACGCCAGTTGGCGCGCGCTCCAGGGGCGCTTCCGCTT 390
115 ArgThrGln.....SerArgGlyArg..... 121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 CGCCACCAACGGTTGGCACCGCTGTCTCCCGAGGCCAAGCTCATCG 440
121 ..... 121
441 CCGAGACACTCCGCGAGTGGCCGAGGAGAGAGTACCCCGTGGAGGTG 490
122 .....GlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgGluGlu 136
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

491 GAGGAGGGGGACCC...CGTGGTGCTGCCCTGTGACCCCGCGAGAGCGC 537
137 ThrGln.....GluLeuArgAlaAlaArgSerArgLeuArgAspPr 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
538 TGTTCCTCCCTAAATCTATTGCTCACAGGACATCTGTTACATCGCTC 587
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 oileLysProGlyMetPheGlyTyrglyValProPheAlaLeuProL 168
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 AGGACGAGAGGTCTCTATGGGCGAGGATGGAACTCTACTTCTCCAAC 637
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 euHisArgAsuArgHisProArgSerProProArgSerGluLeu...S 184
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
638 GCCATGGTGGCGACAG.....CCACCCGACTACATCTGCCA 675
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 exLeuLeuSerSerArgGlyGluGluAlaLeuProSerProThrProArg 200
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
676 CGCTCACTCTCTG.....GCCCGCGACCATCATCCAGA 710
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProTh 217
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 AGGAGCCCTCGACCTCGCGCTGGCCCGCAGTAATCGGTTTCGCTCCGC 760
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 rGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG 234
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
761 CCCCCCGCTGCTGCTGCCCGCGACCCCGACCCCAACGACCATCGCCCT 810
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 luThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArg 250
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
811 CCGGGGG.....CGAGGCTGCTGTGGAGTGCATCGCTGAGG 848
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 HisHisProArgAlaGlnAlaSerGly.....ThrGluProProSe 264
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
849 GCGTCCCGCTCCATCGGTCGATGGCGGCGCTGACGGCCCTCTCTC 898
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 r.....ProThrHisSerLeuGlyGluGlyGlyPheP 275
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
899 CGGGCGGCTTGGAACTTCAACAAACGCTGCGGCTGTGGGGGTGAC 948
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
275 heArgAlaSerPro..... 279
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
949 GAGAGCGAGCGGGAGTACGATGCTGGCTGAGAACGGGAGGGA 998
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
280 GlnProArgArgProSerSerGlnGlyTrpAlaSerProGlnVal..... 294
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
999 CGGCCAGGGGGACCCACAGCTCAGCGTGGAGCGGCGCCCATATTGGGTG 1048
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
295 .AlaGlyArgArgProAspProPheProSerValProArgGlyArgGly 311
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1049 CGCGGCCACAGA.....GTGGGTCTTCGGGCGGGGGAGACGCG 1089
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 lnglnglnglnglnglnglnglnglnglnglnglnglnglnglnglng 327
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 GAGGCTGGACTCGAGGTGGGGGA..... 1115
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
328 LeuGluProAspProGlnHisProGlyAla.....Tr 338
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1116 ...AACCCGACCCCAATCCATGAGCATCAATGGGTCCCATCGAG 1162
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
338 pLeuProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrp..... 352
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1163 GCTGCGCGGGGAGCGGCTGCTGCGGGCGGGCTTGTGTCTTCC 1212
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
353 .....Ser 353
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1213 GGAGCTGCGCCGACAGACAGCGCGGTGCTGAGTCCGAGCGGAGAAC 1262
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 LeuPheAlaProSerSerProIleProArgCys.....SerGlygl 367
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1263 GCCACGGCCCCCTATTGGCCACGCGCTTCTGACGCTCGTGGAGTGC 1312
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 uSerGluGlnLeuArgAlaCysSerGlnAlaPro..... 378
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1313 CTCGGAATGCTGACGGCGGATGACAGCGCTACGAAAGTGTGGAAAAACA 1362
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

379 .....CysProProGluGlnPro..AspProArgAlaLeuGln.. 390  
1363 AACAGTGTTCCTGCAGTGCAGAACCTTCGGGCGCCCGCGCAACAGTGC 1412  
391 .....CysAlaAlaPheAsnSerGlnGluPheMetGlyC 402  
1413 AGTGGCTGACCCCACTTGGAGCGGCTCTGCAGGACACCGATCCTC 1462  
402 lnLeuTyrGlnTrpGluProPheThrGlu..... 411  
1463 GTGTTCAACCAATGGAGCCTTCGCTGAGTGGCTGCGGGGGGGACGG 1512  
411 ..... 411  
1513 GGGGTCTACACCTGCATGTCGCCCAAAACGCCACAGCAACGGCAGCTCA 1562  
412 .....ValGlnGlySerGlnArgCysGluLeuAsnCysArgProArgGlyP 427  
1563 CGCGCTCTCTGGAGGTGAGACGCCCAACCGGAATTCGGGCCCGCCCGC. 1611  
427 heArgPheTyrValArgHisThrGluLysValGlnAspGlyThrLeuCys 443  
1612 .....AAGCGCACCGCGCAAAAGGGAGAGCGGTGACCTTC. 1649  
444 GlnProGlyAlaProAspIleCysValAlaGlyArgCysLeuSerPro... 459  
1650 .....ACTGCGGGGCGACCTTTGACCCCGCC 1675  
460 .....GlyCysAspGlyIleLeuGlySerGlyValArgProA 472  
1676 GTGACCCCGGGGAGTCGATGCTGCGGGGGGGCAGCGCTGCCGA 1725  
472 spGlyCysGlyVal.CysGlyGlyAspAspSerThrCysArgLeuValse 488  
1726 CGACCCCGGTATTTCGTGCGCGGAGATGAC..... 1758  
488 rGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLysIleLeuT 505  
1759 .....GGTGCCAACTGGACTAT 1777  
505 rpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSer 521  
1778 GG.....GGACGAGGGACCAATTCAGTGGCGCGCTCCACCCCTCT 1818  
522 SerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSerIleIleAsnG 538  
1819 .....CGACTCCGCGGAGCGCAAGCGCAGCTCAGAG 1850  
538 yAsnTrpAlaValAspProProGlySerTyrArgAlaGlyGlyThrValP 555  
1851 TCGTGGCGCGCCCGCCATCCCGGACCTCCAAAGTGATGAGGT..... 1893  
555 heArgTyrAsnArgProProArgGluGluGlyLysGlyGluSerLeuSer 571  
1894 .....GGACGAACCGCGT.....GGCCTCAGCTGGAC 1923  
572 AlaGluGlyProThrThrGln.ProValAspValTyrMetIlePheGln 588  
1924 CCCGGGGACGACCATACAGCCCAATAGAGAAGTTCGTGGTGGAGGAG 1973  
588 luGluAsn.....Pro 591  
1974 AGGAGAGAGAGAGATCTTCAGCGGGGTTCGGAGCGGTGAGCTCCG 2023  
592 GlyValPheTyrGlnTyrValIleSerProProIleLeuGluAs 608  
2024 GGGCA.....GCGGTGGACGCCCGCCCTCCCGCTGC..... 2055  
608 nProThrProGluProProValPro..... 616  
2056 .CCCATACGGGGGTTCCTCGTTCGGGTGGCGGTAAACGCTACGGG 2104  
617 .....GlnLeuGlnProGluIleLeuArgValGluProProLeuAla 630  
2105 AGGGGGGAGCACACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 2154  
631 ProAlaProArgProAlaArgThrPro.....GlyThrLeuGlnArgG 645  
2155 TCCGAGCGCAACCGGGGGGTTCATGGGGAGGCAATGAGACCGCA 2204  
645 nValArg..IleProGlnMetProAlaProProHisProArgThrProLe 661  
2205 ACCTGTCATCACCTGGAGCCCTCCCGCCCGCCCGCCCGCCCGCCCG 2254  
661 uGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerA 678  
2255 TGGGCGCGGTACCGG..... 2270  
678 laserCysGlyLysGlyValTrpArgProIlePheLeuCysIleSerArg 694  
2271 .....TGCAGTGGCGCATTTGGAGG.....AGCCCG 2297  
695 GluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaArgPr 711  
2298 GCGGGGGG.....GCCCTTCGGGGGGTTCCTCGTGG 2329  
711 oProAlaSerProGluProCysHisGlyThrProCysProProTyrTrpG 728  
2330 GCGAAAGCACCG.....TGGACGCC...CCCCGTGTGTGT 2364  
728 luAlaGlyGlu..... 731  
2365 GGGGGGGTCCCGCCCGTTCAGCCCTTCAGATCCCGCTCCAGCGCGTGA 2414  
732 .....TrpThrSerCysSe 736  
2415 ACAGGCGGGAAGGACCGGAAGCGACCCCGCGGTGGGGGCACAGCGGG 2464  
736 rArg.....Ser.CysGlyProGlyThrGlnHis 745  
2465 GAGGACCTCGCTGTGTTTACCTTGAGAAATGTGGGGTGGAACTGCTGAA 2514  
746 ArgGlnLeuGlnCysArgGln...GluPheGlyGlyGlySerSerVa 761  
2515 CAGCACACCGTGGCTGAGTGGACTTTGGGGGGGGGC..... 2555  
761 lProProGluArgCysGlyHisLeu..... 769  
2556 ....CAAAGAGCTGGGGGGGTCTGAGGGGTTCGGGTGCTGTATTG 2601  
770 .....ProArgProAsn 773  
2602 GCCTTTGGATGGTGGGGAGCGGAGTCCCGTCAAGCCCGCCCGGACC 2651  
774 IleThrGlnSerCysGlnLeuArgLeu..... 782  
2652 CCCCCCAATCCCAAGCCCGCTGAAGACCCCGCCCGCCCGCCCGCC 2701  
783 .....CysGlyHisTrpGluValGlySerProTrpSerGlnC 795  
2702 GTGGCTCTGACGTGGGGGGGACGCGGGGGGGCGCTGCTGGGGGGGT 2751  
795 ysSer.....ValArgCysGlyArgGlyGlnArgSerArgGln 807  
2752 GCGCCCTGGAGCGGTATTCAGCTGCGGG..... 2780  
808 ValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluCysAl 824  
2781 .....TGTTGGTTCACGGGAGGGGGGACGCGCCCGCCCA.....GC 2818  
824 aserGlyProProGlnProProSerArgGluAlaCysAspMetGlyProc 841  
2819 GAACCCATCGCTTCGAGACCCCGAGGAG...TTCGGGCCCGCCCGA 2865  
841 ysThrThrAlaTrpPheHisSerAspTrpSerSerIysValSerPro 856

```

|||||.....||| |||:::|||||
2866 GGAGCTCCCGGTGG.....ACGGTTGGAACGACACCOCCTCTCCG 2906

seq_name : /cgn2_6/ptodata/2/ins/6A_COMB.seq;US-09-320-878-19

;; sequence information
;; Sequence 19, Application US/09320878A
;; Patent No. 6117659
;; GENERAL INFORMATION:
;; APPLICANT: ASHLEY, Gary
;; APPLICANT: BETLACH, Melanie C.
;; APPLICANT: BETLACH, Mary C.
;; APPLICANT: MCDANIEL, Robert
;; APPLICANT: TANG, Li
;; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;; FILE REFERENCE: 300622002120
;; CURRENT APPLICATION NUMBER: US/09/320, 878A
;; CURRENT FILING DATE: 1999-05-27
;; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;; EARLIER FILING DATE: 1998-08-28
;; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;; EARLIER FILING DATE: 1998-05-06
;; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;; EARLIER FILING DATE: 1997-04-30
;; EARLIER APPLICATION NUMBER: 60/119,139
;; EARLIER FILING DATE: 1999-02-08
;; EARLIER APPLICATION NUMBER: 60/100,880
;; EARLIER FILING DATE: 1998-09-22
;; EARLIER APPLICATION NUMBER: 60/087,080
;; EARLIER FILING DATE: 1998-05-28
;; NUMBER OF SEQ ID NOS: 34
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 19
;; LENGTH: 38506
;; TYPE: DNA
;; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

alignment_scores
      Quality:   345.50           Length:    1077
          Ratio:     0.798         Gaps:        68
Percent Similarity: 40.204       Percent Identity: 25.441

alignment_block
US-10-041-770-2 x US-09-320-878-19 ..

Align seg 1/1 to: US-09-320-878-19 from: 1 to: 38506

      45 GlyValTrpGlyPro...TrpValGlnTrpAlaSerCysSerGlnPro.. 59
      ||||| ||| |||||:::||| |||||:::|
20002 GGGGTGGGCCTCCGGTGGTGCGTAGTCGTGTTCCGGGACAG 20051
      60 .....CysGlyValGlyValGlnArgArGserArGrThrCysGlnLeuP 74
      ||| ||| |||||:::||| |||||:::|
20052 GGCACGAGTGGCGCGGATGGTGGCGAACTCCTCGACGTGTCGAAGGA 20101
      74 roThrValGlnLeuHisProSerLeu..... 82
      ::::: |::|
20102 GTTCGGCGGCCCATGGCCGAGTGCGAGGCGCGTCGTCCTACGTGG 20151
      83 .....ProLeuProToArGroProArghisProGluAl 94
      ||| |||||:::|
20152 ACTGGTCGTGAGGCGGTCTGTCGACAGCCCCCGCGCGC..... 20193
      94 aLeuLeuProArgGlyGlnGlyProArgProGlnThrSerGluThrL 111
      ||||| ::|||
20194 .....CCACGCTGGAGCGGGTGGATG...TGTCCAGCCCGGTGAGTT 20233
      ||| ::|
111 euProLeuTyArGrThrGlnSerArg.GlyArggly.....GlyProLe 125
      ||| ::|
20234 CGCCGTCATGG.....TCTCGTGTGGGAAGGTCTGCGCACGACACAG 20274
```

```
21103 CGCGGAGACCCGCTGCTCGCGGGCGGTGGCGCTCGCGACTCCGAC 21152
|||||
373 ....AlaCysSerGlnAlaProCysProProGlnGlnProAsp..... 385
|||||
21153 GCGTGGCTGCTACGGGAGCTCTCCCTCCGTACGACCCCTGGCTGGC 21202
|||||
386 .....ProArgAlaLeuGlnCysAlaAlaPheAsnSerGln 397
|||||
21203 GGACACAGCGGTGGCGGACACCGTCTGTCGCGGGAACGGGTTTCG... 21249
|||||
398 GluPheMetGlyGlnLeuTyrGlnTrp.....GluProPheThrGlu... 411
|||||
21250 .....TGGAGCTGGCGTTCGAGCGCGGGACACGATC 21281
|||||
412 .....ValGlnGlySerGlnArgCysGlnLeuAsnCysArgp 424
|||||
21282 GGTGCGATCTGTCGAGGAGCTACCCCTCGACGCGCGCTCGTCTGCC 21331
|||||
424 roArgGlyPheArgPheTyrValArgHisThrGluLysValGlnAspGly 440
|||||
21332 CC.....GTCTGGCGGG 21345
|||||
441 ThrLeuCysGlnProGlyAlaProAspIleCysValAlaGlyArgCysLe 457
|||||
21346 TCGGTGTGAGC.....TGTCCGTGGCGCGAGCGAC 21377
|||||
457 uSerProGlyCysAspGlyLeuLeuGlySerGlyArgArgProAspGly 473
|||||
21378 GAGTCCGGCGCTC.....GTACCTTCGGGCTCTACCGCACCCGGAGGA 21421
|||||
474 CysGlyValCysGlyAspSerThrCysArgLeuValSerGlyAs 490
|||||
21422 CGCGCGGGGAGCGAGTGAGCGGGACCGCCCGGTGTGTGGCG 21471
|||||
490 nLeuThrAspArgGlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleP 507
|||||
21472 C.....CCGTGGCGGACCGACCGCCCTCGCCGAC 21503
|||||
507 roAlaGly..... 509
|||||
21504 CGGAGCGCTGGCGCGCGCGGGCGCGAGCGGTTGAGCTGGACGGTCT 21553
|||||
510 .....AlaLeuArgLeuGlnIleAlaGlnLeuArgProSer..... 521
|||||
21554 GTACGAGCGGTTCGCGGCAACGGCTACGGCTACGGCCCGCTTCACGG 21603
|||||
522 .....SerAsnTyrLeuAlaLeuArgGlyP 530
|||||
21604 GCGTCCGTGTGTCTGGCGCGTGGCGGAGGAGTGTTCGCCGACGTGGCC 21653
|||||
530 roGlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspPro...Pro 545
|||||
21654 CTGCGGCGGAGTGCCTGGTGGCGAGGGCGCGGTTCGGCTTCACCC 21703
|||||
546 GlySer.....TyrArgAlaGlyGlyThrVal..... 554
|||||
21704 GCGCTGCTCGACGCGCGCTGTCAGGCGCGCGGTGGCGGGGCGCTTCG 21753
|||||
555 .....P 555
|||||
21754 CGCGGCGACCGCGCTCCGTTCCGCTGGAGCGGATCTCCCTGTACGG 21803
|||||
555 heArgTyrAsnArgProProArgGluGluGly..... 565
|||||
21804 GTGCGGCCACCGCCCTCCCGCTGCGGCTGGCCCGCGCGCGCGACAC 21853
|||||
566 .....LysGlyGluSerLeuSerAlaGluGly..... 574
|||||
21854 GGTGCTCGTGGCGCGCGCACTCTCCGGGACCGGGTGTTCGCCGCGG 21903
|||||
575 ....ProThrThrGlnProValAspValTyrMetIlePheGlnGluGlu 589
|||||
```

```

813 nAenGlyAspGluValSerGluGlnGluCysAlaSerGlyProProGlnP 830
      ::::::::::: |||||
22774 .....GACAGTCCGATCGGATCCGGCCACCGGTCT 22807
830 roProSerArgGluAlaCysAspMetGlyProCysThrAla..... 844
      ||| ::::: |||||
22808 CAACCTCCCGCAGCTCTGATCGCTCGGCTGATCCCGCATGACCCGATCGGGC 22857
      ||| ::::: |||||
845 ...TTPheHisSerAsp.....TTPSerSerLysValSerPro..... 856
      ||| ::::: |||||
22858 TGTATGGCAGCGAGGAGCGCGGTGTACCGGACCGCGCCCGCGCTC 22907
      ||| ::::: |||||
857 .....GluProProAla 860
      ::::::::::|
22908 AGCACCTCGCCCGCGCG 22926

```

seq\_name: /cogn\_2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-103-840A-2

```

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
  Quality: 344.50      Length: 957
  Ratio: 0.946        Gaps: 54
  Percent Similarity: 38.036      Percent Identity: 25.392

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

74 ProThrValGlnLeuHisProSerLeuProLeuProProProAr 90
||||: ||| |||: ||| |||: ||| |||: |||
3927959CGGCGCGCGGTGGCGCGCGCGCGCGCGCTTGGCGCGCTG 3927910

90 gHisProGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThrS 107
||||| ||||| ||||| ||||| |||||
3927909AAGCCGGAGAGCCAGCCGAGACCGTGGCTCCTCGCG...CCGT 3927863

107 erProGluThrLeuPro.LeuTyArgThrGlnSerArgGlyGly 123
||||| ||| |||: ||| |||: ||| |||: |||
3927862CAGCGCACACCGCGGTTCGCCCGCGCGCGCGCTTACCGCG... 3927818

123 yProLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluI 140
||||| ||| |||: ||| |||: ||| |||: |||
3927817.CCCCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3927772

140 leArgAlaAlaArgSerArgLeuArgAspProLysProGlyMet 156
||| ::::::::::|

```

```

3927771CCCGCGGTGGCGCGGAGCGCGCGACGC.....CCAGGG... 3927737
157 PheGlyTyGlyArgValProPheAlaLeuProLeuHisArgAsnArg... 172
      ::||| |||||
3927736.....TTCCGCGACACCGCGCGCGCT 3927714
173 ...ArgHis...ProArgSerProProArgSerGluLeuSerLeuLeSe 187
      |||: ||| |||||
3927713TTGCGCGCGGTGGCGCGGTGGCGCGCGCGCGCGGTGGCGGTGATC... 3927666
187 rSerArgGlyGluGluAlaLeuProSerProThrProArgAlaGluProp 204
      ||| ::::: |||||
3927665.....GAGGTGGTGTGGCGCGCGCGCGCGCGCTTGGCGCGCTT 3927626
204 heSerAlaAsnGlySerProGlnThrGluLeuProProThrGlu...Leu 219
      :: |||||
3927625GG.....CCGCGCGCACCGCTGGCGCGCGCGCTTGAATTTGT 3927588
220 SerValHisThrProSerProGlnAlaGluProLeuSerPro...GluTh 235
      ||| |||||
3927587TCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCGCGCTTGGCGCG 3927538
235 rAlaGlnThrGluValAlaProArgThrArgProAlaPro...LeuArgH 251
      ::||| |||||
3927537GGCGCGCGGTGGCGCGGTGGCGCGGTGGCGCGCGCGCGCGGTGGCG 3927488
251 isHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrHis 267
      ||| ||| ::::: |||||
3927487GGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG... 3927441
268 SerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArgArgPr 284
      ||||| |||||
3927440.....TTGTCGCGCGAGACCTGGCGCAC 3927418
284 oSerSerGlnGlyTrpAlaSerProGlnValAlaGly.ArgArgProAsp 300
      ||||| |||||
3927417GTTGCC.....GCCGATCCCGCGCGCGCGCGCGCGCGCGCGCG 3927377
301 Pro...PheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyPr 316
      ||| |||||
3927376CGGTGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927327
316 oTrpGlyThrGlyThrProHisGlyProArgLeuGluPro..... 330
      ||| ::::: |||||
3927326TTGG...CCGCCGTCAAAGCGCGAGAGCGCGCGCGCGCGCGCGCG 3927280
331 .....AspProGlnHisProGlyAlaTrpLeuProLeuLeu 342
      ||||| |||||
3927279CATCCCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCACTA 3927230
343 SerAsnGlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSe 359
      ::||| |||||
3927229CCGTTGGCGCGCAACG.....CCGCGCGGTGCC 3927201
359 rProfileProArgCysSerGlyGluSerGluGlnLeuArgAlaCysSerG 376
      ||||| |||||
3927200CGGTACCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927151
376 lnAlaProCysProProGluGln.....Pro 384
      ||||| |||||
3927150TGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927101
385 AspPro.ArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMetG 401
      ||||| |||||
3927100GACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3927070
401 lyGlnLeuTyGlnThrGluProPheThrGluValGlnGlySerGlnArg 417
      ::||| |||||
3927069...CGCTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927022
418 GysGluLeuAsnCysArgPro..... 424
      |||||
3927021TGT.....TGCCCGCGCGCTCGCGCTTCTCGCGCGCGCGCGCGCG 3926981

```



GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS: 15  
ADDRESS: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13987 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..13987  
US-08-804-227C-13

alignment\_scores:  
Quality: 342.00 Length: 954  
Ratio: 0.969 Gaps: 42  
Percent Similarity: 37.002 Percent Identity: 22.432

alignment\_block:  
US-10-041-770-2 x US-08-804-227C-13/rev ..  
Align seg 1/1 to reverse of: US-08-804-227C-13 from: 1 to: 13987

80 ProSerLeuProLeuProArgProArgHisProGluAlaLeu 96  
||||| ||| ||||||||| |||||  
10613 CCGGATCCCGCACTCCCGGAGACCCACCCACCCGACACCCGAC 10564  
96 upro ArgGlyGlnGlyProArgProGlnThr.....Ser 107  
||| ||||| ||||| ||||| |||||  
10563 ACCACACACACCGCGCGGACCCGACCCGACCACTCCACACAC 10514  
108 ProGluThrLeuProLeuArgThrGlnSerArgGlyArgGlyPr 124  
||| ||||| ||||| ||||| ||||| |||||  
10513 CCCCAGGAACCTCCCGCACTCCCGGCTCCATCAACAC...CGAATGAAAC 10467  
124 OLeuArg.....GlyProAlaSerHisLeuGlyArgGluThrG 138  
||| ||||| ||||| ||||| |||||  
10466 CATGCGACATCCACACCGCGGACCTCCACCCCATCCCGGCTG 10417  
138 InGluIleArgAlaAlaArgSerArgLeuArgAspProIleLysPro 154  
||| ||||| ||||| ||||| ||||| |||||  
10416 ACACACCGGACTCCGACACACCCACATCCCGGACACACCCGACCG 10367  
155 GlyMetPheGlyTyrGlyArgValProPheAlaLeuProLeuHisArgS 171

10366 .....CGGCCCATTCACCGCG 10350  
171 nArgArgHisProArgSerProProArgSerGluLeuSerLeuLeuSerS 188  
:|||||  
10349 CAACGACACCGCACT..... 10335  
188 erArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPhe 204  
||||| ||||| ||||| |||||  
10334 .....CCCCCAACCCCTCAACAACCCGCC... 10311  
205 SerAlaAsnGlySerProGlnThrGluLeuProThrGluLeuSerVa 221  
:||||| ||||| ||||| |||||  
10310 .....GCACACCGGACTCCGACGCGCGCCACGACCATCCCC 10273  
221 HisThrProSer.....ProGlnAlaGluProLeuSerProGluThrA 236  
||||| ||||| ||||| ||||| ||||| |||||  
10272 CCACCCACCGGACACCCATCAACCCACCCCGCGGCGCACCAACCG 10223  
236 la.....GlnThrGlu 239  
|||  
10222 CACCGCATCACCCAAACGACACACCCCGCCACATACGACGACGACCT 10173  
240 ValAlaProArgThrArgProAlaProLeuArgHisHisProArgAlaG 256  
||||| ||||| ||||| ||||| |||||  
10172 CCCCACCGGATGACCCAAACACCCACCGACACCTCCACACCCCGGACCTCC 10123  
256 nAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGlyG 273  
:||||| ||||| ||||| ||||| |||||  
10122 AACGCGCGGACACGCGCACT...CCAACGCAAAACG..... 10086  
273 lypPhePheArgAlaSerProGlnProArgArgProSerSerGlnGlyTrp 289  
:||||| ||||| ||||| ||||| ||||| |||||  
10085 .....CAGGCTGAGCAAACTCCGTCGACCCCAACACCCACCCCAAAA 10044  
290 AlaSer.....ProGlnValAlaGlyArgArgPr 299  
||| ||||| ||||| ||||| |||||  
10043 CCGAGTCAGCATCCGGAACATCATCCGCAACGACGACGACCATCCACC 9994  
299 oAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyP 316  
| ||| ||||||||| ||||| ||||| |||||  
9993 TCCCCACACCGACACACCTCATCCAGACCTCCGCAACACCCACCCACCC 9944  
316 roTIpGlyThrGlyThrProHisGlyProArgLeuGluProAspPro 332  
||| ||||| ||||| ||||| ||||| |||||  
9943 CCCCAGTACACCCGCGCCCATACCAACCCACTGCGCACCTGAC... 9897  
333 GlnHisProGlyAlaTrpLeuProLeuSerAsnGlyProHisAlaSe 349  
||| ||||| ||||| ||||| ||||| |||||  
9896 .....CCGTGAACACACACCCCGCCACCGACGACCC 9868  
349 rSerLeuTrpSerLeuPheAlaProSerProIleProArgCysSerG 366  
9867 C.....GCACCGCACTCCCGCGCACCA..... 9846  
366 lylGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProGlu 382  
:||||| ||||| ||||| ||||| |||||  
9845 ..CAGCAGCGGCGCGGAGACACACCCACAGGAGCGCGGACGACCA 9798  
383 GlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluP 399  
:||||| ||||| ||||| ||||| ||||| |||||  
9797 CATCCAGACCGCGCACTC.....C 9775  
399 heMetGlyGlnLeuThrGlnTrpGluProPheThrGluValGlnGlySer 415  
:||||| ||||| ||||| ||||| ||||| |||||  
9774 TCACGCGCGCGGCGGACACCGGACGCGCTCAGCA..... 9737  
416 GlnArgCysGluLeuAsnCysArgProArg.....GlyPheArgPh 429  
:||||| ||||| ||||| ||||| ||||| |||||  
9736 CGACGACGCGGACACCAACGACCCCGGACATCCCGGATCCGAC 9687  
429 eTyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro 445  
:||||| ||||| ||||| ||||| ||||| |||||

691	sIleSerArgGluSerGlyGluLeuAspGluArgSer.....	704
692	: : :	
693	: : :	
694	: : :	
695	: : :	
696	: : :	
697	: : :	
698	: : :	
699	: : :	
700	: : :	
701	: : :	
702	: : :	
703	: : :	
704	: : :	
705	: : :	
706	: : :	
707	: : :	
708	: : :	
709	: : :	
710	: : :	
711	: : :	
712	: : :	
713	: : :	
714	: : :	
715	: : :	
716	: : :	
717	: : :	
718	: : :	
719	: : :	
720	: : :	
721	: : :	
722	: : :	
723	: : :	
724	: : :	
725	: : :	
726	: : :	
727	: : :	
728	: : :	
729	: : :	
730	: : :	
731	: : :	
732	: : :	
733	: : :	
734	: : :	
735	: : :	
736	: : :	
737	: : :	
738	: : :	
739	: : :	
740	: : :	
741	: : :	
742	: : :	
743	: : :	
744	: : :	
745	: : :	
746	: : :	
747	: : :	
748	: : :	
749	: : :	
750	: : :	
751	: : :	
752	: : :	
753	: : :	
754	: : :	
755	: : :	
756	: : :	
757	: : :	
758	: : :	
759	: : :	
760	: : :	
761	: : :	
762	: : :	
763	: : :	
764	: : :	
765	: : :	
766	: : :	
767	: : :	
768	: : :	
769	: : :	
770	: : :	
771	: : :	
772	: : :	
773	: : :	
774	: : :	
775	: : :	
776	: : :	
777	: : :	
778	: : :	
779	: : :	
780	: : :	
781	: : :	
782	: : :	
783	: : :	
784	: : :	
785	: : :	
786	: : :	
787	: : :	
788	: : :	
789	: : :	
790	: : :	
791	: : :	
792	: : :	
793	: : :	
794	: : :	
795	: : :	
796	: : :	
797	: : :	
798	: : :	
799	: : :	
800	: : :	
801	: : :	
802	: : :	
803	: : :	
804	: : :	
805	: : :	
806	: : :	
807	: : :	
808	: : :	
809	: : :	
810	: : :	
811	: : :	
812	: : :	
813	: : :	
814	: : :	
815	: : :	
816	: : :	
817	: : :	
818	: : :	
819	: : :	
820	: : :	
821	: : :	
822	: : :	
823	: : :	
824	: : :	
825	: : :	
826	: : :	
827	: : :	
828	: : :	
829	: : :	
830	: : :	
831	: : :	
832	: : :	
833	: : :	
834	: : :	
835	: : :	
836	: : :	
837	: : :	
838	: : :	
839	: : :	
840	: : :	
841	: : :	

```
seq_documentation_block:
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
```



APPLICANT: Rosteck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

alignment\_scores:  
Quality: 342.00 Length: 954  
Ratio: 0.969 Gaps: 42  
Percent Similarity: 37.002 Percent Identity: 22.432  
alignment\_block:  
US-10-041-770-2 x US-08-804-227C-7/rev ..  
Align seg 1/1 to reverse of: US-08-804-227C-7 from: 1 to: 44377  
80 ProSerLeuProLeuProArgProArgHisProGluAlaLeuLe 96  
||||: ||| |||||: ||||| |||||  
10628 CCGATCCCCCACTCCCGGACCCACCCACCCCGGACCGGAC 10579  
||| ||| |||||: |||||  
96 uProArgGlyGlnGlyProArgProGlnThr.....Ser 107  
||| ||| |||||: |||||  
10578 ACCACCACACACCGCGCGCACCGGACCTCCACGACTCCACAC 10529  
108 ProGluThrLeuProLeuArgThrGlnSerArgGlyGlyPr 124  
||| ||||| |||||: |||||  
10528 CCCCCGGAACCTCCCCACACCGCGGCTCCATCAAC...CGAATGAAC 10482

124 oLeuArg.....GlyProAlaSerHisLeuGlyArgGluGluThrG 138  
||: ||| ||||| ||| :  
10481 CATGGACACATCAACCGCGGCGACTCCACCCATCCCGATCAGCGTG 10432  
138 InGluIleArgAlaAlaArgArgSerArgLeuArgAspProIleLysPro 154  
||: ||| ||||| ||||| :  
10431 ACACACCGGACTCCAGCACACCCACATCAACCCGACACACACCGGACCG 10382  
155 GlyMetPheGlyTyrglyArgValProPheAlaLeuProLeuHisArgAs 171  
||: ||| ||||| ||||| :  
10381 .....CGGCCCATTCACCGCG 10365  
171 nArgArgHisProArgSerProProArgSerGluLeuSerLeuIleSer 188  
||: ||| ||||| ||||| :  
10364 CAACCGACACCCACT..... 10350  
188 erArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPhe 204  
||: ||| ||||| ||||| :  
10349 .....CCCCAACCCCTCAACACCCCC... 10326  
205 SerAlaasnGlySerProGlnThrGluLeuProProThrGluLeuSerVa 221  
||: ||| ||||| ||||| :  
10325 .....GCACCAACCGACTCCGACGCGCCCGCCACCGACCATCC 10288  
221 lHisThrProSer.....ProGlnAlaGluProLeuSerProGluThrA 236  
||: ||| ||||| ||||| :  
10287 CCACCCCGGACACCCACCGATCAACCCCGCGCGCCACCGACCG 10238  
236 la.....GlnThrGlu 239  
||: ||| ||||| ||||| :  
10237 CACCGCATCAACCAACGACACACCCCGCCACATACGACGACGACCT 10188  
240 ValAlaProArgThrArgProAlaProLeuArgHisHisProArgAlaG 256  
||: ||| ||||| ||||| :  
10187 CCCCCCGAATGACCAACACGACCGACCTCCACACCCCGGACGCTCC 10138  
256 nAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluGlyG 273  
||: ||| ||||| ||||| :  
10137 AACGCGCGGAACAACGCGACCT...CCAACGCAAAACAACG..... 10101  
273 lYPhePheArgAlaSerProGlnProArgArgProSerSerGlnGlyTrp 289  
||: ||| ||||| ||||| :  
10100 .....CAGGCTGAGCAAACTCCGCGACCCACCAACACCCCAAAA 10059  
290 AlaSer.....ProGlnValAlaGlyArgArgPr 299  
||: ||| ||||| ||||| :  
10058 CCGAGTCAGCATCCGGAACATCACTCCCGCAACGACCGACCATCCACC 10009  
299 oAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyP 316  
||: ||| ||||| ||||| :  
10008 TCCCCACCCCGACACACCTCATCCAGCACCTCCCGCAAAACACCCACCC 9959  
316 roTrpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspPro 332  
||: ||| ||||| ||||| :  
9958 CCGCATACACACCCCGCGCCATACCAACCCACTGGCGACCTGAC... 9912  
333 GlnHisProGlyAlaTrpLeuProLeuSerAsnGlyProHisAlaSe 349  
||: ||| ||||| ||||| :  
9911 .....CCGTGAACAACACCCCGCCCGCCACCGACCC 9883  
349 rSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerG 366  
9882 C.....GCACCGCATCCCGCGCACCA..... 9861  
366 lYgUSeRGlUlnLeuArgAlaCysSerGlnAlaProCysProProGlu 382  
||: ||| ||||| ||||| :  
9860 ..CAGCACCGGCGCGCGAAGACACACCCACAGGAGCGCGCGACCGACCA 9813  
383 GlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluP 399  
||: ||| ||||| ||||| :  
9812 CATCCAGACCGCGCGCAACTC.....C 9790  
399 heMetGlyGlnLeuTyrglnTrpGluProPheThrGluValGlnGlySer 415

9090 TTGTCGCGACCGTCTCTGATTGACCGACCTGCCCCCGACCCAGCCAGCAC 9041  
683 ..... GlyValTyrArg. .... 686  
9040 CGGATACCGTTGCGCGCGCGCTCGGAAAGCGCTCCACACGACGACAC 8991  
687 ..... ProIlePheLeuCy 691  
8990 CCACACCGCTCCGCCACGCGCTCCGCTCGGACACCTCCGCGAACCGCTTG 8941  
691 stleSerArgGluSerGlyGluLeuAspGluArgSer. .... 704  
8940 CACCGCGCGCTCCGCCACGCGCTCCGCTCGGACGAGAACTCCACGAAAC 8891  
704 ..... 704  
8890 CGTCGGCGCTGCCATCACGTCACCCACCGCCACCGCCACGACGATCACACT 8841  
705 ..... CysAlaAlaGlyAlaArgProProAlaSerProGluProCysHis 719  
8840 CACCGCGCGCAACGACTGCACCGCGAGATCAACGCGCCACCAACGACGAC 8791  
719 sGlyThrProCysProProTyrTrpGluAlaGlyGluThrSer. .... 734  
8790 GAACACGCGTGTCCACCGCTCAGCGCGCGCTCCACCCCAACGACGATA 8741  
734 ..... 734  
8740 GGCACACCGCGGAGATCACACTGCGGAGGTCCTCCCGTGAGCACATGCC 8691  
735 ..... CysSerArgSerCysGlyProGlyThrGlnHisArgGlnLe 748  
8690 CGCGTAGGTGTGCGCGCGCTCCCGAGCTGGTGCCTGCTGCTGCTGCTG 8641  
748 uGlnCysArgGlnGluPheGlyGlyGlySerSerValProGluA 765  
8640 GAGATCGCGTGAAGACACGCGTCCGGGTCCCGCGAGGAGTCCGGAAC 8591  
765 rgCysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArg 781  
8590 GATCGCGCGCTCCAGCGCTCCACGACGCTCTCCACGACCAACCGCT 8541  
782 LeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSerVal. 797  
8540 GCTCGCGGTCCATGG. .... CCGTGGCTCGCGCGGCGATATC 8503  
797 ..... 797  
8502 CCGAAGAAGCTCCGCTGAACTCGGCGCGGAGAACGACGAGAACCGGCTTC 8453  
797 ..... 797  
8452 CGGTACATAGCTCTTCGCGGACACCGCGCTCCGGGTCTGACAGCCCT 8403  
798 ..... ArgCysGlyArgGlyGlnArgSerArgGlnVa 808  
8402 CGACGTCACGACGCTCGCGCGGAGACGCGGAGAGCGTGTCCCGTCA 8353  
808 lArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluCysAla 825  
8352 CCGCGCACACGAGTCCACAGTCTCGCGGACGCGACCGACCGCGGGTA 8303  
825 erGlyProProGlnProProSerArgGluAlaCysAspMetGlyProCys 841  
8302 ACGGACGCGCATGCCACGATCGGATCGGCTCACGCTCGTCGGCCA... 8256  
842 ThrThrAlaTrpPheHisSerAspTrpSerSerLysValSerProGluPr 858  
8255 TGACCGCGCTGG. .... GCGCGCGCGCTCGCGGCTCC 8224  
858 oPro 859  
8223 GCCT 8220

seq\_name: /cgn\_2\_6/ptodata/2/ina/5b\_comb.seq:US-08-804-198-1

seq\_documentation\_block:  
; Sequence 1, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rosteck, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; STREET: PAUL R. CANTRELL 1138  
; CITY: LILLY CORPORATE CENTER  
; STATE: INDIANAPOLIS  
; COUNTRY: IN  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 31329..36071  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 36155..41830  
; US-08-804-198-1

alignment\_scores:  
Quality: 342.00 Length: 954  
Ratio: 0.969 Gaps: 42  
Percent Similarity: 37.002 Percent Identity: 22.432

alignment\_block:  
US-10-041-770-2 x US-08-804-198-1/rev ..

Align seg 1/1 to reverse of: US-08-804-198-1 from: 1 to: 44377

80 ProSerLeuProLeuProArgProArgHisProGluAlaLeuLe 96  
||||: ||| |||||:||||| |||||

10628 CCGATCCCCCAACTCCCGGAACCCACCACCCACCCGACACACCCGAC 10579  
96 uProArgGlyGlnGlyProArgProGlnThr .....Ser 107  
||| :|||:||||  
10578 ACCACACACACCGCGCCGACCCGACCGAACTCCCAAGACTCCACAAC 10529  
||| :|||:||||  
108 ProGluThrLeuProLeuTyArgThrGlnSerArgGlyArgGlyPr 124  
||| ||||| :|||:||||  
10528 CCCCAGGAACCTCCCAACACCGGCTCCATCAACAC...CGAATGAAC 10482  
||| ||||| :|||:||||  
124 cLeuArg.....GlyProAlaSerHisLeuGlyArgGluGluThrG 138  
|||:||||| ||||| ||| :|||:||||  
10481 CATCGACACATCCACCGCGGCACTCCACCCATCCCCCATCAGCGTG 10432  
|||:||||| ||||| ||| :|||:||||  
138 lngluLeuArgAlaAlaArgSerArgLeuArgAspProLysPro 154  
:||||| ||| :|||:||||  
10431 ACAACACCGACTCCAGCACACCCACATCAACCCGACACACCGACCG 10382  
|||:||||| ||||| ||| :|||:||||  
155 GlyMetPheGlyTyGlyArgValProPheAlaLeuProLeuHisArg 171  
|||:||||| ||||| ||| :|||:||||  
10381 .....CGGCCGATTCACCGCG 10365  
171 nArgArgHisProArgSerProProArgSerGluLeuSerLeuLeuSer 188  
:||||| ||||| ||| :|||:||||  
10364 CAACCGACACCCACT..... 10350  
188 erArgGlyGluGluAlaLeuProSerProThrProArgAlaGluProPhe 204  
|||:||||| ||||| ||| :|||:||||  
10349 .....CCCCCAACCCCTCAACACCCCC... 10326  
205 SerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSerVa 221  
|||:||||| ||||| ||| :|||:||||  
10325 .....GCACCGGACTCCGAGCGCCCGACCGACCATCC 10288  
221 lHisThrProSer.....ProGlnAlaGluProLeuSerProGluThrA 236  
||| ||||| :|||:||||  
10287 CCACCGACCGGCAACCCACCATCAACCCCGCGCGCCACCAACCG 10238  
||| ||||| :|||:||||  
236 la.....GlnThrGlu 239  
||| ||||| :|||:||||  
10237 CACCGCATCCCAAGGACAACACCCCGCCACATACGACGACGACCT 10188  
||| ||||| :|||:||||  
240 ValAlaProArgThrArgProAlaProLeuArgHisHisProArgAlaG 256  
|||:||||| ||||| ||| :|||:||||  
10187 CCCCACCGAATGATCCCAACACCGACACTCCACACCCCGGAGCTCC 10138  
||| ||||| :|||:||||  
256 nAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluGly 273  
:||| ||||| :|||:||||  
10137 AACGCCGGAACAGCGCCACT...CCAACGCAACACACG..... 10101  
||| ||||| :|||:||||  
273 lypPheArgAlaSerProGlnProArgArgProSerSerGlnGlyTrp 289  
:||||| ||| :|||:||||  
10100 .....CAGGCTGAGCAAACTCCGTCGACCCCAACACCCCAACAAA 10059  
||| ||||| :|||:||||  
290 AlaSer.....ProGlnValAlaGlyArgArgPr 299  
||| ||||| :|||:||||  
10058 CCGATCAGCATCCCGGAACATCATCCCGACGACCGACCATCCACC 10009  
||| ||||| :|||:||||  
299 oAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGly 316  
||| ||||| :|||:||||  
10008 TCCCCACCGACGACACACACTCATCCAGCACTCCGCAAAACACCCACC 9959  
||| ||||| :|||:||||  
316 rotPrglyThrGlyThrProHisGlyProArgLeuGluProAspPro 332  
||| ||||| :|||:||||  
9958 CCCCAGTACAAACCCGCGCCCATCAACCCACTGCGCACCCTGAC... 9912  
||| ||||| :|||:||||  
333 GlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSe 349  
||| ||||| :|||:||||  
9911 .....CCGTGAACACACACCCCGACCGACGCC 9883  
||| ||||| :|||:||||  
349 rSerLeuTrpSerLeuPheAlaProSerSerProLysProArgCysSerG 366  
||| ||||| :|||:||||  
9882 C.....GCACCGCATCCCGCGCACCA..... 9861  
||| ||||| :|||:||||

366 lyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProProGlu 382  
 9860 ..CAGCAGCGGCCCCGAGACACACACAGGAGCCCCGCGACCA 9813  
 383 GlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluP 399  
 9812 CATCAGACCGCGCAGCACTC.....C 9790  
 399 heMetGlyGlnLeuTyrGlnTrpGluProPheThrGluValGlnGlySer 415  
 9789 TCACGCCCGCGCCACCCACCGCAGCGCTCACCGAA..... 9752  
 416 GlnArgCysGluLeuAsnCysArgProArg.....GlyPheArgPh 429  
 9751 CGAGCAGCGGACACACCAACAGCCCGCCACATCCGCGGATCCGCAC 9702  
 429 eTyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro 445  
 9701 CCACCGGTCACACA.....CTCAGCAACCGCACCGCTCCGCCCGC 9658  
 446 GlyAlaProAspIleCysValAlaArgCysLeuSerProGlyCysAs 462  
 9657 AACGCCCGC.....TCGACCGCGCAGAAAGCACCCACGACAA 9620  
 462 pGly...IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysG 478  
 9619 CGGACCGCGTCACCGGTCGGCGGCTCAGGTACAGGCGCGTCCCGCTGCT 9570  
 478 lyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArg 494  
 9569 CG..... 9568  
 495 GlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLe 511  
 9567 .....CTCCGCGGCTCCGCC..... 9553  
 511 uArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuA 528  
 9553 ..... 9553  
 528 rGlyProGlyArgSerIleIleAsnGlyAsnTrpAlaValAspPro 544  
 9552 ..GGCGCTCTCCAGGACC.....ACATGGCGGTGGTCCCG 9517  
 545 .....ProGlySerTyrArgAlaGlyGlyThrValPheArgTyrAsnAr 559  
 9516 CTCACCGGAAAGCC.....GACACCGCGCGCGCGCCAGCCCG 9479  
 559 g.ProProArgGluGluGlyLysGlyGluSerLeuSerAlaGluGlyPro 575  
 9478 CTCACCGCGCGCGCCACGACCGGG...TCTCGGTACAGCTCCACCG 9432  
 576 ThrThrGlnProValAspValTyrMetIlePheGlnGluGluAsnProGl 592  
 9431 CACCCGAAGCCACTCCACTTCGACGAGCGGCATPCACATGCAGCGGTC 9382  
 592 yValPheTyrGlnTyrValIleSerSerProProIleLeuLeuAsnP 609  
 9381 CGGGCAACAGCCATGCGGCATCCCTGCACCACTTGATCACCACACC 9332  
 609 roThrProGluProProVal...ProGlnLeuGlnProGluIleLeuArg 624  
 9331 CACACCGGACGCGCGCTGGCATGCCGCTGTCGACTTCAACAGCCGCA 9282  
 625 ValGluProProLeuAlaProAlaProArg.....ProAlaArgTh 638  
 9281 GCACAAAGGATCGCGACCGCTCCCGCGGTACGTGGCCAGCAGCGCA 9232  
 638 rProGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProP 655  
 9231 CCGCGCTCGATCGGATCACCCAGCGCGCTCCCGTATCCGTCGCCCTCCAC 9182

```
8302  ACGGCGCATGCCACGATCGGATCGCTACGCTCGTGGGCA... 8256
842  ThrAlaThrPheHisSerAspTrpSerLysValSerProGluPr 858
8255  TGACCGCGGTGG.....GGCGGCGCGCTCGCGGCTCC 8224
858  OPR 859
8223  GCCT 8220
```

seq\_name: /sgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-105-537-32

```
seq_documentation_block:
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-32
```

alignment\_scores:  
Quality: 341.50 Length: 1212  
Ratio: 0.759 Gaps: 64  
Percent Similarity: 37.129 Percent Identity: 22.937

alignment\_block:

US-10-041-770-2 x US-09-105-537-32 ..

Align seg 1/1 to: US-09-105-537-32 from: 1 to: 11220

```
46 ValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysGlyVa 62
623  TTGTGGAGCGACTGGCGGGTTCGACGGGCTGCTCCGTCGCGG..... 6564
62  lGlyValGlnArgSerArgThrCysGlnLeuProThrValGlnLeuH 79
6564 .....
79  lProSerLeuProLeuProPro.....ArgProProArg 90
6565 ..CGGTCAACGGGCTTACCGCCACCGGTTCGGGCGGACCCAGCA 6712
91  HisProGluAlaLeu.....LeuProArgGly.....GlnGlyPr 102
6713  TCCAAGAGCTCGTCAGCGGTGTGAGCCGCGGTCGCGCACGATC 6762
102  oArgProGlnThrSerProGluThrLeuProLeuTyArg..... 115
6763  ATCCCGCTGACTACGGCTCCACAGCGCCACGTCGAGACCATCGAGAG 6812
116  ....ThrGlnSerArgGlyArgGlyPro.....LeuArgGly 127
6813  CGAACTCGCGGACGCTCTGGGGGGTGTCCCGCCAGACACCCAGGTCC 6862
128  ProAlaSer.HisLeuGlyArgGluGluThrGlnGluLeuAlaAla 144
6863  CCTTCTTCTCACACCTCGAAGCGGCTTGATCATCCGAACCCGCTCGAC 6912
144  rArgSerArgLeuArgAspPro.....IleLysProGly 155
```

```
6913  GCGGCTACTGTTACGCGCAACCTCGGCATCGTGTGGGCTTCGCCCCGGC 6962
156  MetPheGlyTyrglyArgValProPheAlaLeuProLeuHisArgAsnAr 172
6963  CGTGAACCCCTGGCCACCGAGAGGTTTACCCACTTCTGTCGAGGTCA 7012
172  gArgHisProArg..SerPro.....ProArgSe 181
7013  GCGCCACCCCGTCTCTACCATGGCGTGGCCGAGACCGTCCCGGACTC 7062
181  rGluLeuSerLeuIleSerSerArgGlyGluGluAlaIleProSerProT 198
7063  GGCACCCCTCCCGGTGACACGGGAGACACCGCTTCACCACTCCCT 7112
198  hrProArgAlaGluProPheSerAlaAsnGlySerProGlnThr..... 212
7113  CGCCGAGGCTGGGCA.....ACGGCTCAGCGTCAGTGGGCT 7153
213  .....GluLeuProThrGluLeuSerValHisThrProSe 225
7154  CTCCTCTCCCGACACCGACCCACCGCATCTGCCACCTACGCTTC 7203
225  r...ProGlnAlaGluProLeuSerProGluThrAlaGlnThrGluVala 241
7204  CAGACCGAGCGCTACTGGCCGCGAGCCG.....ACCTCTCGC 7241
241  laProArgThrArgProAlaProLeuArgHisHisPro.....ArgAla 255
7242  CGCCGCTGACATCACCTCCGCGGTCTCGGGCGCGGACCGCGTGC 7291
256  GlnAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGlu 272
7292  TCGGCGCGCGCTGGCGTCCGGGACTCCGAGCGCTGCTGCTCAGGGG 7341
272  yGlyPhePheArgAlaSerProGlnProArgArgProSer.....SerG 287
7342  A.....GCTCTCCCTCGTACGACCCCTGGCTGGCGGACCA 7379
287  lnglyTrpAlaSerPro.GlnValAlaGlyArgArgProAspProPhePr 303
7380  CGCGGTGGCGGACCGTGTCTGCTCCGCG..... 7409
303  oSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProTrp..... 317
7410  .AACGGCGTTCGTGGAGCTGGGTTCCGAGCGGGGACGAGTGGTTGC 7458
318  ....GlyThrGlyThrPro.....HisGlyProArgLeuGluProAsp 331
7459  GATCTGGTCGAGGAGTCACTCCCTCAGCGCGCTGCTGCTGCCCGCTG 7508
332  ProGlnHisProGlyValaTrpLeuProLeuLeuSerAsnGlyProHisAl 348
7509  TGGCGGCTCCGTCGTGTGTCAGCTGTCTCGGCGGAGACGAGTCCGGC 7558
348  aSerSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCys 365
7559  GTCTGACCTTCGGGCTCTACGCGCACCGGAGGACGCGCGGCGCA...G 7605
365  exGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProPro 381
7606  GCGGAGTGACGCGGACCGCACCGGTGTGTGCGCGCGCTGCGGACCG 7655
382  GluGlnProAspProArgAlaLeuGlnCysAlaAlaPhe..... 394
7656  CACCGCCCGCTCGCGCACCGGAGCGCTGCGCGCGCGGCGCGGAGC 7705
395  .....AsnSerGlnGluPheMetGlyGln..... 402
7706  CGGTGGACGTGGACGCTCTGACGAGCGCTTCGCGCGGACCGCTACGCG 7755
403  .....LeuTyrglnTrpGluPro 408
```



```
772 .....ProAsnIleThrGlnSerCysGln 779
9492 CTGCTCGCCGCGAGTTCGTGCGACCCCTCGCTGC.....GCTGCTCG 9535
780 LeuArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSe 796
9536 GGCCGCGCGCCGCTGCTGAGATGGGAGAACCCAGCAGCTCC.....GC 9579
796 rValArgCysGly..... 800
9580 GACCGGAGCGGTCCCGCGCGACCCCGGTGTCGGCTACCGCGCCTT 9629
801 .....ArgGlyGlnArgSerArgGlnValArgCys..... 810
9630 CGACCTGGCGAGCGCGCGCGAGCGGATCGCGAGATGCTCGCGAGG 9679
811 .....ValGlyAsnAsnGlyAspGluValSerGluG1 821
9680 TCATCGCCCTTCGAGGAGGGGTGCTCCGGACCTCGCCGCTCAGACCC 9729
821 nGluCysAlaSerGlyProGlnPro.....ProS 832
9730 TGGGAGCTGGCGCGCGCGCGACGCTTCGGACAGTCAGCGAGCCCG 9779
832 erArgGluAlaCysAspMetGlyProCysThrThra..... 844
9780 CCACAGGGCAAGGTGCTCTCAGATGCGTGGCGGCTCGACCCGAGG 9829
845 ..... 854
9830 GTACGGTCTGTCACCGCGCGACCGGTGCGTGGGGGCGATCGTGGCC 9879
846 eHisSerAspTrpSerSerIysVal..... 854
9880 CGCACCTGTGTGGCGAGTGGGCGCTACGACGCTCTGCTGTGAGCCG 9929
855 .....SerProGluProAlaIleSerCysIleLeuGlyAsn 867
9930 GCGGGGACGAGCCCGCGCGCGCGAGCTCGTGACGAGCTGGAGG 9979
868 HisAlaGlnAspThrSerAlaPhePro 876
9980 CCCTGGGAGCGAGCTCGGTGGCGCG 10006

seq_name: /cgn_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-5
seq_documentation_block:
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

alignment_scores:
Quality: 341.50 Length: 1212
Ratio: 0.759 Gaps: 64
Percent Similarity: 37.129 Percent Identity: 22.937

alignment_block:
US-10-041-770-2 x US-09-105-537-5
Align seg 1/1 to: US-09-105-537-5 from: 1 to: 36778
46 ValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysGlyVa 62
22310 TTGTGGACGACTGGCCGGGTTCGACGGGCTGTCGCTGCGCG..... 22351
62 lGlyValGlnArgArgSerArgThrCysGlnLeuProThrValGlnLeuH 79
22351 ..... 22351
79 lSProSerLeuProLeuProPro.....ArgProProArg 90
22352 ..CCGTCACCGGGCTACCGCACCGTGTTCGGGCGCACCCGACCCAGA 22399
91 HisProGluAlaLeu.....LeuProArgGly.....GlnGlyPr 102
22400 TCCAGAGCTGCTCAGCGGTGTGAGGCCGAGGGGTCCGGCGACGGATC 22449
102 oArgProGlnThrSerProGluThrLeuProLeuTyrArg..... 115
22450 ATCCCGCTCGACTACGCTCCACAGCGCCACGTCGAGACCATCGAGAG 22499
116 .....ThrGlnSerArgGlyArgGlyPro.....LeuArgGly 127
22500 CGAACTCCGCGACGTCCTGGCGGGGTTCCTCCCGACAGACACCCAGGTCC 22549
128 ProAlaSer.HisLeuGlyArgGluGluThrGlnGluLeuAlaAla 144
22550 CTTCTTCTCCACCTCGAAGCGGCTGGATCAGCGAACCCGCGCTCGAC 22599
144 rGArgSerArgLeuArgAspPro.....IleIysProGly 155
22600 GCGGCTACTGTGTCACCAACCTCGCCATCGTGTGGGCTTCGCCCCCGC 22649
156 MetPheGlyTyrGlyArgValProPheAlaLeuProLeuHisArgAsnAr 172
22650 CGTGAAACCTTGCGCCACCGAGAGGCTTCACCCACTTCGTCGAGGTCA 22699
172 gArgHisProArg...SerPro.....ProArgSe 181
22700 GGGCCACCGCTCTCTCACCATGGCGCTGCCCGAGACCGTCACCGGACTC 22749
181 rGluLeuSerLeuIleSerArgGlyGluGluAlaIleProSerProT 198
22750 GGCACCTTCGCGCTGACACCGCGGACAGCACCGCTCACCACCTCCCT 22799
198 hrProArgAlaGluProPheSerAlaAsnGlySerProGlnThr..... 212
22800 CGCCGAGGCTGGGCA.....ACGGCTCACCCTCGACTGGGCT 22840
213 .....GluLeuProProThrGluLeuSerValHisThrProSe 225
22841 CTCCTCTCCCGACGACCGACCCACCGGATCTGCCACCTACGCTTC 22890
225 r...ProGlnAlaGluProLeuSerProGluThrAlaGlnThrGluVala 241
22891 CAGACGAGCGCTACTTGGCGCGAGCGCG.....ACCTCTCGC 22928
241 laProArgThrArgProAlaProLeuArgHisHisPro.....ArgAla 255
22929 CGCCGGTGACATCACCTCCGCGGCTCGGGGCGCCGAGCACCGCTGC 22978
256 GlnAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluG1 272
22979 TCGGCGCGCGCTGGCGCTCGGGACTCGACGGCTGCTGCACGGG 23028
272 yGlyPhePheArgAlaSerProGlnProArgProSer.....SerG 287
23029 A.....GCTCTCTCTCTGTCAGCACCCCTCGCGGAGCA 23066
287 lnglyTrpAlaSerPro.GlnValAlaGlyArgArgProAspProPhePr 303
```

```

23067 CGGGTGGCGGACCGTGTGTCGCGG..... 23096
2303 oSerValProArgGlyArgGlyGlnGlnGlyProTrp..... 317
23097 .ACGGCGTTCGTGGAGCTGGCTTCGAGCCGGGACAGGTGCGTTCC 23145
318 .....GlyThrGlyGlyThrPro...HisGlyProArgLeuGluProAsp 331
23146 GATCTGTCTGAGGAGTCACTTCGACGGCGCTGCTGCTGCGCGTCG 23195
332 ProGlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAl 348
23196 TGGCGGCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23245
348 aSerSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCys 365
23246 GTCGTACCTTCGGGCTCTACGCGCACCGGAGGACGCGCGGCGA...G 23292
365 erGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProPro 381
23293 GCGGAGTGGACGGGACGCGCACCGCGGTGCTGCTGCGCGCGTGGACCG 23342
382 GluGlnProAspProArgAlaLeuGlnCysAlaAlaPhe..... 394
23343 CACGCGCCCGTCCGCGACCGCGGAGCTGCGCGCGCGCGCGCGAGC 23392
395 .....AsnSerGlnGluPheMetGlyGln..... 402
23393 CGGTGGAGCTGGAGCTCTGTACGAGCGCTTCGCGCGACGCGTACGGC 23442
403 .....LeuTyrlGlnTrpGluPro 408
23443 TACGGCCCCCTTCAGAGGCGTCCGTGGTGTCTGGCGGCGTGGCGACGA 23492
409 PheThrGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProAr 425
23493 GGTGTTCGCGACGTGCGCTTCGCGCGGAGTTCGCGCGTGGCGAGGCG 23542
425 gGlyPheArgPheTyrrValargHisThrGluLysValGlnAspGlyThrL 442
23543 GCGGTTGCGGCTTCACCGCGCGTGTCTGACGCGCGCGCGTGGCGGCGC 23592
442 eucysGlnProGlyAlaProAspIleCysValAlaGlyArgCysLeuSer 458
23593 GGTGCGGCGCGGCGTTCG.....GCGGCGGACGCGGCTGCGGTTCC 23636
459 ProGlyCysAspGlyIleLeu.....G1 466
23637 CCTGGAGCGGATCTCTGTACGCGGTGCGCGCACCGCGCTCGCGTGC 23686
466 ySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAspAspSert 483
23687 GCGTGGCCCCCGCGCGCGGACACGCTGTC.....CGTGAGC 23724
483 hrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGly 499
23725 GCGCGGACTCTCCGCGGACCGGCTGTTCGCGCGGACTCCCTCACGGT 23774
500 TyrlGlnLysIle..... 503
23775 GCTGCGCGTCAACCGCGAGCTGGCGGCTTCAGGACCGGACTGCG 23824
504 .LeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgp 520
23825 ACGCGCTGCACGTGTGGAGTGGACCGCGCTGGGACGCGTGGCGGACGCC 23874
520 roSerSerAsnTyrlLeuAlaLeuArgGlyProGlyArgSerIleIle 536
23875 CT.....GCCGCGCGGCTGCTGCTGGGCGCGGACGCCGA 23909
537 AsnGly.....AsnTrpAlaValAspPro.....Pr 545
: : : : :

```

```

23910 CGGTCTCCCGCGCGCTGCGCGCGGTGGCACCGAGGTCTCTCTCTCC 23959
545 oGlySerTyrrArgAlaGlyGlyThrValPheArgTyrrAsnArgProProA 562
23960 CGGACCTTACGAGACCTGTGGAGG.....CGTCGACCGCGGCG 23997
562 rgGluGluGlyLysGlyGluSerLeuSerAla..... 572
23998 GAGACCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24047
573 .GluGlyPro..... 575
24048 CGATGGCGCGGAGCATGTCCGCGAGGCGCTGCACGGGTGCTGCGGCTGA 24097
575 ..... 575
24098 TCCAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24147
576 .....ThrTh 577
24148 GTGACCGCGGCGGTGCTGCGCGCGCTTCCGCGACGCGCTGCGGTCCAC 24197
577 rGlnProValaspValtyrMetIle.....PheGlnGluGluAsn 591
24198 GGGACAGCGCGCGCTGCTGCGCGCTGCGCGCGTCCGCGGACGAGAGGCC 24247
591 roGlyValPheTyrrGlnTyrrValIle.Ser..... 600
24248 CGGCG.....CGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24288
601 .....SerProProPro..... 604
24289 AGGCGCGGCGGCGCGCGCGCGCGCGCTGACGACGCGGCGGCGCGCC 24338
605 .....IleLeuGluAsnPro..... 609
24339 CGTGGCGCGACCTCTGGAGACGCGCGCTGCGCGCGCGCTGCGCGCGG 24388
609 ..... 609
24389 CCCTGCGCTGCGGCGAGCGCGAGCTGCGCTCCGCGGACGCGGCGCTCTC 24438
610 .....ThrProGluProValProGlnLeuGlnp 620
24439 GTACCGCGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 24488
620 roGlu.....IleLeuArgValGluProProLeuAlaPro 631
24489 CGGCGCGGCGCGCTGCGCGCTGCTGCGCGTGC...CCGCGCGCTCGCGCG 24535
632 Ala..... 632
24536 TCTGGCGTGTGAGCGCGGTGAGCGGCGCGCTGAGAGCGCTCAGCGCG 24585
633 .....ProArgProAlaArgThrProGlyThrLeuGlnArgG 645
24586 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24629
645 lnValArgIleProGlnMetProAlaProPro..... 655
24630 GGGACAGTCCGATCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24679
656 .....HisProAr 658
24680 TCTGTGCGCTGCGCGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCG 24729
658 gThrProLeuGlySerPro...AlaAlaTyrrTrpLysArg..... 670
24730 GGAGCGCGCGTGTGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24779
671 .....ValGlyHisSerAlaCysSer...AlaSerCysGlyLysGlyVal 684
24780 CGGCGCGCGGTCTATGGCGCTGCTCTCCGCGCGGTACGCGCGCGGTCTG 24829

```



```
685 TrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAs 701
|||||
24830 TGGCGG.....ACGCGGGACC 24846
701 pGluArgSerCysAlaAlaGlyAlaArgProAlaSerProGluProC 718
|||||
24847 GTCCGCGGATGCCGAGGGTGGAGCTTCGCCAGGGGCGCT...CGT 24893
718 yHis.....GlyThrProCys..... 723
|||||
24894 GCGGGGTGTTCTCTACGCGCGTCTACGCCCTGGCGGACCTGGCGGACG 24943
724 .....ProTyrTrpGluAl 729
|||||
24944 TCAAGCCGCGGAGCGCTCTCTGGTCCACTCCGCGCGCGGTG.....GC 24987
729 aglyGluTrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisA 746
|||||
24988 GTGGCATGGCGCGCGCTGCAGC.....TCGCCGCGCACTGGGCGT 25028
746 rgGlnLeuGlnCysArgGlnGluPheGlyGly..... 757
|||||
25029 GGAGTCCACGCGCGCGAGTACCGGAGTGGGACGCCCTGGCGGCGC 25078
758 .....GlySerValProGluArgCysGlyHisLe 769
|||||
25079 TCGGCTGGACGCGCGCATCGCTCTCCGCGCACCTGGACTTCGAG 25128
769 uProArg..... 771
|||||
25129 TCGCGGTTCGTCCGCTTCCGCGGGGGGATGGAGTGTACTGAA 25178
772 .....ProAsnIleThrGlnSerCysGln 779
|||||
25179 CTCGCTGCGCGGAGTGTCTCAGCCCTCGCTGC.....GCCGCTCG 25222
780 LeuArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSe 796
|||||
25223 GCGCGCGCGCGGTTCGTGGAGATGGGAAGACCGACCTCC.....GC 25266
796 rValArgCysGly..... 800
|||||
25267 GACCGGAGCGGTTCGCGCGGACCCCGGTCTGGTACCGCGCCTT 25315
801 .....ArgGlyGlnArgSerArgGlnValArgCys..... 810
|||||
25317 CGACCTGGCGGAGCGCGCGCGGAGTGGCGAGATCTCGCGCGAGG 25366
811 .....ValGlyAsnAsnGlyAspGluValSerGluG1 821
|||||
25367 TCATCGCCCTCTCGAGAGCGGGTGTCTCGGCACTGCGCCGTCACGACC 25416
821 nGluCysAlaSerGlyProGlnPro.....ProS 832
|||||
25417 TGGGACGTGGCGCGCGCGCGCTTCCGCGACGTCAGCAGCGCGCG 25466
832 eArgGluAlaCysAspMetGlyProCysThrAla..... 844
|||||
25467 CCACAGGCGCAAGTCTCTCTCAGATGCGGTCTGGGCTTCGACCGCGAGG 25516
845 .....TrpPh 846
|||||
25517 GTACGCTCTCTCTACCGCGGCGCGCGTGGGTGGGGCATCGTGCC 25566
846 eHisSerAspTrpSerSerLysVal..... 854
|||||
25567 CGGACGTGTGTGGGAGTGGGGGTACGACGCTGTCTGTCTGTGAGCCG 25616
855 .....SerProGluProAlaIleSerCysIleLeuGlyAsn 867
|||||
25617 GCGGGGACGAGCGCGCGCGCGGAGCTGTCTCAGAGCTGTGACAGCTGGAGG 25666
```

```
868 HisAlaGlnAspThrSerAlaPhePro 876
|||||
25667 CCCTGGGAGCGGACGTCTCGTGCGCG 25693
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11
```

```
seq_documentation_block:
; Sequence 11: Application US/08458568A
; Patent No: 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
US-08-458-568A-11
```

```
alignment_scores:
Quality: 341.00 Length: 1023
Ratio: 0.881 Gaps: 58
Percent Similarity: 37.830 Percent Identity: 24.340
```

```
alignment_block:
```

```
US-10-041-770-2 x US-08-458-568A-11/rev ..
```

```
Align seg 1/1 to reverse of: US-08-458-568A-11 from: 1 to: 12001
```

```
74 ProThrValGlnLeuHisProSerLeuProLeu...ProProArgProPr 89
|||||
5859 CCACAGCGCGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGGAGAG 5810
89 oArgHis.....ProGluAlaLeu.LeuPro...ArgGlyGln 100
|||||
```

5809	TGCTACGCCCGTGGCGCCATCGGCGTCACTGACCTGTCGGA	CTCG 5766
101	GlyProAArgProGlnThrSerProGluThrLeuProLeuTy	ArgThrGI 117
5759	GGTCTCTCGTCTCGTCTCGGCGCATCTTCTCGTCTCGT	CGCAGG 5710
117	nSerArgGlyArgGlyGlyProLeuArgGlyProAlaSer	HisLeuGlyA 134
5709	GACGAGGACGACGCGCAACACGCGG...CCGACACG	ACGCGGAGCC 5663
134	rgGluGluThrGlnGluLeuAlaAlaArgArgSerArg	LeuArgasp 150
5662	CGGGCGCGTGGCGGGTCCGTCAGCGCGGCGC.....	CGGCGAG 5622
151	ProLeuLysProGly.....MetPheGlyTy	rgGlyArgValPr 163
5621	CCC.....CGGGCGGACGCGGCCCGCGCGGCCAC	CCCCCCTCTCC 5578
163	oPheAlaLeuProLeuHisArgAsnArgArgHisPro	ArgSerProProA 180
5577	GAGCGCGCGCCAAAGCCCGCGGCGGAGACCCCGCG	CGCTCCGCG 5528
180	rgSerGluLeuSerLeuLeuSerSerArgGlyGluGlu	AlaLeuProSer 196
5527	GGCGCGCATCGAC.....CGCGCGGCGCGCGCGG	TGCGCGCCG 5484
197	ProThrProArgAla.....GluProPheSerAla	asnGlySerProGI 211
5483	GCGACGCGACGCGCGCTCAAGCCCGGCGACGCC	CGCGCGGTG 5434
211	nThrGluLeuProProThrGluLeuSerValHisThr	ProSerProGlnA 228
5433	GACGCGGACGCGACCT.....CGCGCGCTCTCTA	5405
228	laGluProLeuSerProGluThrAlaGlnThrGlu	ValAlaProArgThr 244
5404	CGCGCGCTATCGCAGCGGTACGTACGCGGAGCC	GTGCCCGGCGCG 5355
245	ArgProAlaPro.....LeuArgHis	ProArgAl 255
5354	GGCCCCCGCGCGCGGTGTGTACGCGCGCTT	GCGGACGCCGCG 5305
255	aglnAlaSerGlyThrGluProProSerProThr	HisSerLeuGlyGluG 272
5304	CGGGCGCTCTGGGGGCGCGGAGCGGAGCGCG	CGCGGT.... 5259
272	lyGlyPheArgAlaSerProGln...ProArgArg	ProSerSerGln 287
5258	.....TCGAGGCTCGGGCGCGCGCGCGCGTGT	GGGCGC 5223
288	...GlyTrpAlaSerProGln.ValAlaGlyArg	ArgProAspPropheP 303
5222	CCGAGCTGGGCGACCGCGCAGCAGTA.....CG	CCCTGTATCAGCGGG 5179
303	roSerValProArgGlyArgGlyGlnGly.....	..... 313
5178	CTGCTGTACACCCCGGCGGAGCGCATGGG	GGTGTCCAGAACCGCG 5129
314	GlnGlyProTrpGlyThrGlyGlyThrProHis	.....G1 325
5128	CGTGTCCCGGGACGTGGCGCTGGACAGCCTG	TGTTCCGATCTCGG 5079
325	Y...ProArgLeuGluProAspProGlnHisPro	GlyAlaTrpLeuProLe 341
5078	GGCGCGCGCGCAACAGCAGCTCTTCATCAC	CGCGACGCTGG.... 5037
341	uLeuSerAsnGlyPro.....	.....H 347
5036	.....CGCGGCGCTGCCCCACTGGCTAC	CGCATGGCGCGCGCGCG 4994
347	laAlaSerSerLeuTrpSerLeuPheAlaProSer	SerProIleProArg 363
4993	CTTCAGCTGGGCTCTGGCGCACGCGGGGCGG	.....CGTGGCCTAGA 4950

[illegible]

[illegible]

```

3452 GTCCCGCAGGCGGGCGGTCTCTCTGCTGTGTGGGGCGCTTGCCACACC 3400
      |||||::|||||
829   .....GlnProPserArGluAlaCysAspMetGl 839
      :::|||||:::|||
3402 GGCTGTGGGGCGGCACAGCGCGCTGGCGGCAATTGGACCGCGCG 3353
839 vProcysThrThraLtpPheHis.....S 848
      |||||
3352 CCGGACGTCGCGCTGGCGCACAGGCGTCTGCTGTGCCACGCG 3303
      |||||
848 eRasTrpSerLys.....ValSerProGlu 857
      :|||
3302 GGAACCTGGCTTCGCGGGCGCGTGAGTTCTGGGCTGCTGCCAGCG 3253
858 ProProAlaIleSerCysIleLeuGlyAsnHisAlaGlnAspThrSerAl 874
      |||||
3252 CCGGGCACC.....GGCGCTCATCTGTCGTCACACCGTG 3218
874 apHeProAla 877
      :|||
3217 CGCGCGCTGCG 3208

seq_name:/cgn2_6/ptodata/2/ina/5A_COMB.seq;US-08-458-568A-11

seq_documentation_block:
; Sequence 11. Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; TITLE OF INVENTION: Compositions and Methods for Treatment
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus type 1
; STRAIN: Herpes Simplex Virus Type 1

```

[illegible]

|||||  
2737 CGGCTCGC.....GCCCCAGCCACG 2759  
576 rThrGlnProValaspValrMetilePheGlnGluAsnProGlyV 593  
|||||  
2760 TACAGCGCGCGCA.....CGCGCGCGCCACGCGC 2788  
593 alPheYrGlnTyValIleSerSerProProIleLeuGluAsnPro 609  
2789 CCAGCGCGCGCGCGCGCGAGTGGCCCTCTCTCGCAGAGTCCG 2838  
610 ThrProGluProValProGlnLeuProGluIleLeuArgValG1 626  
2839 GCAGCGCGCGCGCGCGCGTGGTGGTCCCGAGG.....CC 2876  
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2877 GCAGCGCGCGCGCGCGCG.....CCGCGCAGCAGCGCGCGGTACTC 2920  
643 lnArgGlnValArgIleProGlnMetProAlaProHisProArgThr 659  
2921 GCAGCGCGCGCGCGCGCGCGTGGTGGTCCG..... 2949  
660 ProLeuGlySerProAlaAlaTyTrpLysArgValGlyHisSerAlaCy 676  
2950 .....GCCCCAGCGCGGTG 2963  
676 s.....SerAlaSerCysGlyLysGlyValTrpArgProIlePheL 690  
2964 CGCAGCGCGGTAGCGACGTGGCGCGCGCG..... 2994  
690 euCysIleSerArgGluSerGlyGluGluLeuAspGluArgSerCysAla 706  
2995 .....ACAGCGCGCAGCGCGCGCGCGTGGGTGACAGCG 3027  
707 AlaGlyAlaArgProProAlaSer..... 714  
3028 GCAGGTGGCGCGCTCCACCGCGCGAGACCCCGCGCGCGGCGGACACGCGG 3077  
715 .....ProGluProCysHisGlyThrProCysProProTyTrpGluA 729  
3078 CCGAGCGCGCAGCACCTGGCGCGCGCGCGTCCCGCGCGCG.....G 3118  
729 lagGlyGluTrpThrSerCysSerArgSerCysGly.....ProGly... 742  
3119 CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3168  
743 .....ThrGlnHisArgGlnLe 748  
3169 CGTGTCTCCCGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3218  
748 uGlnCys.....ArgGlnGluPheGlyGlyGlyG 758  
3219 ACGGTGTGACCGAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268  
758 lySerSerValProProGluArgCysGlyHisLeuProArgProAsnIle 774  
3269 CAGAACTCCACGCGCGCGCGCA.....AGCGCAGGTCCG 3303  
775 ThrClnSerCysGlnLeuArgLeuCysGly...HisTrpGluVal Gly 789  
3304 CGGTGGACACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3353  
790 SerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerAr 806  
3354 CGCGCGGTCCAATTCCCGCGCGCGCGCGCGCGTGTCCGCGCGCGCGCGCG 3403  
806 gGlnValArgCysValGlnAsnAsnGlyAspGluValSerGluGlnGluC 823  
3404 GTTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3453  
823 ysAlaSerGlyProProGln.....ProProSerArgGluAlaCysAsp 837  
.....

3454 ACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3503  
838 MetGlyProCysThrThrAlaTrpPheHisSerAspTrpSerSer 852  
3504 ACGCGCGAGTACAGCACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3553  
853 .....LysValSerProGlu..... 857  
3554 GCGCATCCAGCGCGCGCATGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCG 3603  
858 .....ProProAla 860  
3604 CAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3623

Q9nsj8 homo sapien  
Q9w493 drosophila  
Q9ufz4 homo sapien  
P90884 caenorhabdi  
Q9vfe1 drosophila  
Q9wkv3 drosophila  
Q9xsv8 bos taurus  
Q9v9q7 drosophila  
Q9nrm2 homo sapien  
Q9cr7 mus musculu  
Q95875 homo sapien  
Q96qc6 homo sapien  
Q9lyp4 mus musculu  
Q9hsb6 homo sapien  
Q22631 caenorhabdi  
Q9cx59 mus musculu  
Q9uli7 homo sapien  
Q9h8x0 homo sapien  
Q9zlr1 mus musculu  
O41935 murid herpe  
Q9veg7 drosophila  
Q9uq1 homo sapien  
Q9ttes5 bos taurus  
O61603 drosophila  
Q9p3j0 neurospora  
Q9bxj3 homo sapien  
Q90584 gallus gall  
Q99307 epstein-bar

17 466 9.5 550 4 Q9NSJ8  
18 460.5 9.4 1054 5 Q9W493  
19 414.5 8.5 988 4 Q9UFZ4  
20 403.5 8.2 1059 5 P90884  
21 391.5 8.0 1229 5 Q9VF61  
22 357 7.3 566 5 Q9WKV3  
23 351.5 7.2 1637 6 Q9XSV8  
24 346.5 7.1 4123 4 Q9V9Q7  
25 325.5 6.6 858 5 Q9V9Q7  
26 318 6.5 432 4 Q9NPM2  
27 305.5 6.2 244 11 Q9CRC7  
28 301 6.1 2157 4 Q95875  
29 301 6.1 2157 4 Q96QC6  
30 296 6.0 235 11 Q9LYP4  
31 292.5 6.0 237 4 Q9HSB6  
32 284 5.8 788 5 Q22631  
33 282.5 5.8 192 11 Q9CX59  
34 281.5 5.8 1023 4 Q9ULI7  
35 268.5 5.5 822 4 Q9H8X0  
36 268 5.5 2157 11 Q9ZLR1  
37 264.5 5.4 585 12 O41935  
38 264 5.4 2703 5 Q9VEG7  
39 262 5.4 364 4 Q9UCQ1  
40 260.5 5.3 685 6 Q9RTS5  
41 260 5.3 2715 5 O61603  
42 255 5.2 2592 3 Q9P3J0  
43 251 5.1 417 4 Q9BXJ3  
44 251 5.1 1146 13 Q90584  
45 250.5 5.1 924 12 Q99307

# ALIGNMENTS

RESULT 1  
ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.  
AC Q9EPX2  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE PAPILIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1;  
RX MEDLINE=20530499; PubMed=11076767;  
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,  
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,  
RA Sieron A.L., Prockop D.J., Fessler J.H.;  
RT "Papilin in development; a pericellular protein with a homology to the  
RT ADAMTS metalloproteinases";  
RL Development 127:5475-5485(2000).  
DR EMBL; AF314171; AAG41980.1; -;  
DR HSSP; P12111; 2KNT.  
DR InterPro; IPR003599; Ig.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003600; Ig-like.  
DR InterPro; IPR003006; Ig\_MMC.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR Pfam; PF00047; Ig; 3.  
DR Pfam; PF00014; Kunitz\_BPTI; 1.  
DR Pfam; PF00090; tsp.1; 5.  
DR PRINTS; PR00759; BASICPTASE.  
DR SMART; SM00409; IG; 3.  
DR SMART; SM00408; IGc2; 3.  
DR SMART; SM00410; IG-like; 2.  
DR SMART; SM00131; KU; 1.  
DR SMART; SM00209; TSPI; 5.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: July 23, 2002, 21:04:46 ; Search time 67.92 seconds  
(without alignments)  
2233.754 Million cell updates/sec

Title: US-10-041-770-2  
Perfect score: 4895  
Sequence: 1 MENWGRPWLYLLLSLPQ.....PPAISCILGNHAQDTSAPFA 877  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19.\*  
1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mbc.\*  
8: sp.organelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.rodent.\*  
12: sp.virus.\*  
13: sp.vertibrate.\*  
14: sp.unclassified.\*  
15: sp.virus.\*  
16: sp.bacteriap.\*  
17: sp.archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	738.5	15.1	1280	11 Q9EPX2	Q9epx2 mus musculu
2	731	14.9	3198	5 Q9D8G8	Q9d8g8 manduca sex
3	728.5	14.9	1235	4 Q95428	Q95428 homo sapien
4	714	14.6	2174	5 Q9GQR0	Q9gqr0 drosophila
5	711.5	14.5	3060	5 Q9YAV4	Q9yav4 drosophila
6	695.5	14.2	1572	5 O44938	O44938 haemochus
7	692.5	13.5	454	4 Q96H81	Q96h81 homo sapien
8	661.5	13.5	538	4 Q9H8E4	Q9h8e4 homo sapien
9	640.5	13.1	2167	5 Q78840	Q78840 caenorhabdi
10	618	12.6	951	4 O60345	O60345 homo sapien
11	618	12.6	2165	5 Q19791	Q19791 caenorhabdi
12	532.5	10.9	1427	4 Q96L37	Q96l37 homo sapien
13	530	10.8	525	4 Q96RW4	Q96rw4 homo sapien
14	510.5	10.4	766	4 P82987	P82987 homo sapien
15	505	10.3	1014	5 Q95R33	Q95r33 drosophila
16	477.5	9.8	340	11 Q91Z56	Q91z56 mus musculu

```
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.  
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.  
DR PROSITE; PS00092; TSPI; 3.  
KW Immunoglobulin domain; Serine protease inhibitor.  
SQ SEQUENCE 1280 AA; 138824 MW; AE287705B561AF30 CRC64;  
  
Query Match 15.1%; Score 738.5; DB 11; Length 1280;  
Best Local Similarity 31.6%; Pred. NO. 3.3e-43;  
Matches 173; Conservative 50; Mismatches 226; Indels 99; Gaps 13;  
  
QY 335 PCAWPLLSNGPHASSLWSLAPSPPIR-----CSGSEQL 371  
DB 16 PSW---ARNVRQSDTWGTWGESPCSTCGGSIREFPCYSQRDRDGTSCVGPASRH 72  
  
QY 372 RACSAQPCPEQDPDRAALQCAAFNSQEFMGQLYOWEPTEVOGSORCELNCRPRGFYV 431  
DB 73 RYCHTESCPDGVDRFRAEQCAEFQDFTQGRYRWLPY--YAAPNKCELNCPKQGFY 130  
  
QY 432 RHYEVQGTLCQPCAPDICVAGCLSPGCGILSGRRPDCGCGGDDSTCLVSNL 491  
DB 131 KHKDAVDGTCEPQQRDICVDGCRVVGCDHKLDISKQEDKLCQCGDGGSCYPVTGTF 190  
  
QY 492 TDRGGPLGYQKLTIPAGALRLQIAQLRPSNLYALRGFGGRSIINGNWAYDPGSRAG 551  
DB 191 DGNLDSRGVQNIIFIIPAGATSIRIEEAAARNFLAVKSIRGEYILNGHWTIEAAQALPVA 250  
  
QY 552 GVTFYRNPRPREGKESLSAEGTPTQVDVYMFIOENPGVYQYVVISPPPLENTP 611  
DB 251 STVLYERGVGDLAPERLQARGPTSEPLVIELLSQESNPGVHYEYL----- 298  
  
QY 612 EPPVPQLQPEILLRVPPLAPAPRPARPTGTQLQVRIPQMPAPPHRPTPLGSPAAYNKR 671  
DB 299 -----PANDPCR-----GFSWSHG 312  
  
QY 672 GHSASACGKGWRPIFLICISRESGEELDSRCAAGARPASPEPCHGTCP--PYEA 729  
DB 313 SMGDCSAECGGHGRQLVECTI--DNEAYPDHMCQHPR--PTHRSCNTQPCPKTKRWKV 369  
  
QY 730 GEWTSRSCGPGTQHRQLQCRQFEGGSSVPPE--RGCHLP--RPNITQSCOLRLCGHWE 787  
DB 370 GWPVPCSVSCGGVQSRVYCLSSDGTGGQEAEEQACAGLAGKPTTQACNLQCAVWS 429  
  
QY 788 VGSFWSQSVRCGRGORSRQVRCVGNNGDEVSEQECASGPPPPSREACDMGPTT---- 843  
DB 430 V-EPMGECSTCGTGIRKRSVTCRDEGSPVHAAACLL-KDQPTLTPEVCQBPVFRG 487  
  
QY 844 AMFHSWMS 851  
DB 488 AMHVGWS 495  
  
RESULT 2  
Q9UBG8 PRELIMINARY; PRT; 3198 AA.  
AC Q9UBG8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE LACUNIN PRECURSOR.  
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Spingilodea; Spingilidae; Spinginae; Manduca.  
OX NCBI_TaxId=7130;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99457716; PubMed=10528409;  
RA Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;  
RT "Expression of lacunin, a large multidomain extracellular matrix  
protein, accompanies morphogenesis of epithelial monolayers in Manduca  
sexta.";  
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
```

```
DR HSSP; P12111; 2KNT.  
DR InterPro; IPR004094; Antistatin.  
DR InterPro; IPR003598; Ig_C2.  
DR InterPro; IPR003006; Ig_MHC.  
DR InterPro; IPR002223; Kunitz_BPTI.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR002221; WAP.  
DR Pfam; PF02822; Antistatin; 4.  
DR Pfam; PF00047; Ig; 2.  
DR Pfam; PF00014; Kunitz_BPTI; 9.  
DR Pfam; PF00095; wap; 1.  
DR PRINTS; PR00759; BASICPTASE.  
DR SMART; SM00131; KU; 10.  
DR SMART; SM00209; TSPI; 7.  
DR SMART; SM00217; WAP; 1.  
DR PROSITE; PS00317; 4-DISULFIDE_CORE; 1.  
DR PROSITE; PS00280; BPTI_KUNITZ_1; 8.  
DR PROSITE; PS00279; BPTI_KUNITZ_2; 10.  
DR PROSITE; PS00092; TSPI; 1.  
KW Immunoglobulin domain; Serine protease inhibitor; Signal.  
FT SIGNAL 1 21 POTENTIAL.  
SQ SEQUENCE 3198 AA; 349364 MW; AB4ACD459C0D9134 CRC64;
```

Query Match 14.9%; Score 731; DB 5; Length 3198;  
Best Local Similarity 28.5%; Pred. NO. 3e-42;  
Matches 179; Conservative 73; Mismatches 216; Indels 160; Gaps 17;

```
QY 305 VPRGRGQGGQGWGT-----GGTPHCPRLPEPDPQHPGAWLPILLSNGPHASSLWSLF 355  
DB 55 IPGEGE--TGWDGMDSTPCSRTCGGVASQKRI-----CLKFGPDQ----- 95  
  
QY 356 APSPIPRCSGESQLRACQAPCPPEPQPPRALQCAAFNSQEFMGQLYOWEPTEVQGS 415  
DB 96 -----PQCTGGDTKYFSCETQDCPGSTDFRABQCSKYDDVEYRNKTKKPYT--RGP 147  
  
QY 416 QRCLENCRPRGFYRHTKVDGTLCCQPCADPICVAGCLSPGCGIILGSRPRPGCG 475  
DB 148 NPCELNCMPHGERYYRHQLKVIDGTRCNDDSDFYCVNGTCQPVGCCDMLGNSAREDKCR 207  
  
QY 476 VCGDDSTCLRVSNLDRGGPLGYKILMIPAGALRLQIAQLRPSNLYALRGPPGGRSI 535  
DB 208 KCRGNKCYNTNVLTQDLIKYNDMLLIPGATTIVIMEVKPSNNYLALRAKNITYY 267  
  
QY 536 INGNWAYDPGSGYRAGTVFYRNPRPREKGSLSAEGTPTQPDVYMFIOENPGVY 595  
DB 268 LNGDYHIDFPRSMIAGALWFYERSQQGFAAPDKRLCLGPTTEPLXLSLLQSYNVGIEY 327  
  
QY 596 QYVSISSPPPLENTPPEPPVQLQPEILLRVEPPLAPAPRPARPTGTQLQVRIPQMPAPP 655  
DB 328 EYSL-----PTERAPRTQN----- 342  
  
QY 656 HPRTPLGSPAAYNKRVGHSACSGKGVWRPIFLICISRESGEELDSRCAAGARPASP 715  
DB 343 -----YTWVHEHTECSATCGGGGFQTRNVNCTQMDLELVDDTLCDGLK--PVTN 391  
  
QY 716 EPCHGTPCPYPWEAGETWTSRSG--PCTQHRQLQCRQFEGGSSVPPER--CGHL--PR 771  
DB 392 QTCNNNVCPQMWVTGEWCSHRGCSGDSGTQREIQCKIIVNGFRSIVSKRECDLLGPK 451  
  
QY 772 PNITQSCOLRL--CGHWEYGVSPWQSVRCGRGQSRQVRCVGNNGDEV----- 818  
DB 452 PEKFRKCNVNTCTPTWTG--PWKPCDLYGEGKQTRQVVCVQKNRRRIDVLDGSECTDER 510  
  
QY 819 --SQEC-----ASGPPPPS 832  
DB 511 PESEQKCMHSEERTDWVASEWSGDCNCFSTMRTRIAKTYDRKLVDSFCAHPLPVL 570  
  
QY 833 REACD---MGPCTTAFHSDWSSKVSPE 857  
DB 571 QEPCCAAITLPPCDVQWYATQW--SKCSSE 597
```

QY	708	GARPPASPECHGTGCP	-----YWEAGWTSCRSRSGP	743
Db	346	QPR-PADRRSCNLRHCPETKRTSYLHRPGAWRLAGAVCGNSWKAGPWAPCASCGGS	404	
QY	744	QHRQLQCRQEGGG-GSSVPPERCGLP-RPNITQSCQLRGLGHVWEGSPWSCSVRCGR	801	
Db	405	QSRSVTCISSDGIQAEVBEACAGLPKPPAIAQCNLQRCAAWS-PEFWGECVSCGV	463	
QY	802	QGRSROVRCVGNNGDEVSEQECASGPPQPPREACDMGQC----TTAWFHSW	850	
Db	464	GVKRSVTCGERGSLHLTAAC-SLEDRPPLTEPCVHEDCPLLSDQAWHVGW	515	
RESULT	4			
Q9GORO				
ID	Q9GORO	PRELIMINARY;	PRT;	2174 AA.
AC	Q9GORO			
DT	01-MAR-2001	(Tremblrel. 16, Created)		
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)		
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)		
DE	EXTRACELLULAR MATRIX PROTEIN PAPILIN PRECURSOR.			
GN	PPN.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-DP CN BW;			
RC	MEDLINE-20530499; PubMed=11076767;			
RA	Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,			
RA	Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,			
RA	Sieron A.L., Prockop D.J., Fessler J.H.,			
RT	"Papilin in development; a pericellular protein with a homology to the			
RT	ADAMTS metalloproteinases."			
RL	Development 127:5475-5485(2000).			
DR	EMBL; AF205357; AAG37995.1;			
DR	HSSP; P12111; 2KNT			
DR	Flybase; FBgn0003137; Ppn.			
DR	InterPro; IPR003599; Ig.			
DR	InterPro; IPR003598; Ig_c2.			
DR	InterPro; IPR003600; Ig-like.			
DR	InterPro; IPR002223; Kunitz_BPTI.			
DR	InterPro; IPR000884; TSP1.			
DR	InterPro; IPR002221; WAP.			
DR	Pfam; PF00014; Kunitz_BPTI; 3.			
DR	PRINTS; PR00003; 4DISULPHCORE.			
DR	PRINTS; PR00759; BASICPTASE.			
DR	SMART; SM00409; IG; 3.			
DR	SMART; SM00408; IGC2; 2.			
DR	SMART; SM00410; IG_Like; 1.			
DR	SMART; SM00131; KU; 3.			
DR	SMART; SM00209; TSP1; 7.			
DR	SMART; SM00217; WAP; 1.			
DR	PROSITE; PS50279; BPTI_KUNITZ_2; 3.			
DR	PROSITE; PS50092; TSP1; 3.			
KW	Matrix protein; Serine protease inhibitor; Signal.			
FT	SIGNAL 1 26			
FT	POTENTIAL.			
SQ	SEQUENCE 2174 AA; 231936 MW; 038F707952623120 CRC64;			
Query Match	14.6%;	Score 714;	DB 5;	Length 2174;
Best Local Similarity	29.7%;	Pred. No. 3e-41;		
Matches 174;	Conservative 85;	Mismatches 209;	Indels 118;	Gaps 21;
QY	302	FFSVPRGRGQGGGW- GTGCTPHGPRLEPDHPQHGAWLLPSNGPHASSLWLFAPS-	358	
Db	28	FFGL-RQKQYGANMYLPSSVTPGEGNDP	68	
QY	359	-----SPIPR-----CSGESQLRACSOAPCPPOPPRALQCAAFNSQ	397	
Db	69	CSRTCGGVSYQTRCLRRDRGEAVCGSGSRRTSCNTQDCPEESDFRAQCQSRFRDQ	128	



QY 398 EFMGLYQWFFTEVOGSCRELNCRGRFFRYHTEKVDGTLQCPGAPDVCVAGRCL 457  
DB 129 QFDGVYFVWVPTNAPN-CELNCMPGGERFYRQREKVVGDGRKNDKLDVCGNECM 186  
QY 458 SPGCDGLGSRRRDPGCGVCGDDSTCLVSGNLTDRGGPLGYQKILWIPAGALRLQIAQ 517  
DB 167 PVGCDMLGSDAKEDKCRKCGDGGSTCKTIRNTITKNLAPGYNDLLLPBEGATNIRIEE 246  
QY 518 LPPSSNYLALRGPGRSSTINGNWDVPPGSRAGTGVFRYRNPREECKGSLSAEGPTT 577  
DB 247 TVPSSNYLACRNSHGHYLLNGDWRIDFRPMPFFANSWNYORKPMGFAAPQLTCSGPIS 306  
QY 578 QPVDYVMIQFQENPGVYQYVYSSPPPLENPTBPPVQLOPEILRVEPPLAPAPRPAR 637  
DB 307 ESLFVLMVQEKNSLDYEYSI-----PE-----SLSHSQDQDTH 340  
QY 638 TPGTLQVRIPQMPAPPHPRTPLGSPAAYKRVGHSCASCSGCGVWRPFLICISRESG 697  
DB 341 T-----WTHQFNACASCGGSGSONRKYTCNNRITL 371  
QY 698 BELDERSCAAGARPASPEPCHTGTPCPYWEAGWTSCRSRG-PGTQHRQLQCRQEFEG 756  
DB 372 AEVNSLQDQSK-PVEQACGTETECAPHWVEGWSKCGSGDGFQNRITCERISS 430  
QY 757 GSSVPE-----RCGLHPRPNITQSC--QLKLGHWVEGSPWOCVRCGRGQSRQVR 809  
DB 431 GEHTVEEDAVCLKEVGN--KPATKQECNRDVKNCPKYHLG-PWTPCDKLCGDGKQTKRVT 487  
QY 810 C-VGNNGDF--VSEOCASGPPQPSREACDMGPCT-TAWFHSWDS 851  
DB 488 CFIENGHRKRVLEEDCVE--EKPTKESCLTTCGVDYWIISQWS 531  
RESULT 5  
QYAV4 ID QYAV4 PRELIMINARY; PRT; 3060 AA.  
AC QYAV4: QYAV3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CG1540 PROTEIN.  
GN PPN OR CG1540 OR CG18436.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotlier P.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Center A., Chandra I.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Caudy A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A.S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Warkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mout S.M., Mov M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,  
RA Shue B.C., Siden-Kianos I., Stapleton M., Strong R., Sun E.,  
RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Svirska R., Tector C., Turner R., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC EMBL; AE003765; AAF56794.2; -;  
DR EMBL; AE003765; AAF56795.2; -;  
DR HSP; P12111; 2KNT.  
DR FlyBase; FBgn003137; Ppn.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR003598; Ig\_C2.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR002221; WAP.  
DR Pfam; PF00047; Ig; 3.  
DR Pfam; PF00014; Kunitz\_BPTI; 12.  
DR Pfam; PF00090; tsp\_1; 5.  
DR Pfam; PF00095; wap; 1.  
DR PRINTS; PR00759; BASIPTASE.  
DR SMART; SM00408; IGC2; 3.  
DR SMART; SM00131; KU; 12.  
DR SMART; SM00209; TSPI; 7.  
DR SMART; SM00217; WAP; 1.  
DR PROSITE; PS00317; 4\_DISULFIDE\_CORE; 1.  
DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 11.  
DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 12.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS00092; TSPI; 3.  
KW Alternative splicing; Immunoglobulin domain;  
KW Serine protease inhibitor.  
FT VARSPPLIC 2803 2803 L -> SVVPV (IN SHORT ISOFORM).  
FT VARSPPLIC 2844 2854 FNEKTMDSGI -> VASPLHFNAPV (IN SHORT ISOFORM).  
FT VARSPPLIC 2855 3060 MISSING (IN SHORT ISOFORM).  
SQ SEQUENCE 3060 AA; 331579 MW; ACA31D3EE558C7C0 CRC64;  
Query Match 14.5%; Score 711.5; DB 5; Length 3060;  
Best Local Similarity 31.4%; Pred. No. 6.4e-41;  
Matches 157; Conservative 78; Mismatches 190; Indels 75; Gaps 14;  
QY 364 CSGSEQLRACQAPCPPEQDPDRALQCAANSQEFMQLQWEPFTEVQGSQRCNLNCR 423  
DB 55 CSGGRRYFSCNTQCPPEESDFRAQCSRFDRQDFGVFEWVPYTNAPN--CELNCM 112  
QY 424 PRGRFYRHTKVDGTLQCPGAPDVCVAGRCLSPGCDGILGSGRRDPGCGVCGDDST 483  
DB 113 PKGERFYRQREKVVGDGRKNDKLDVCGNECMVPGCDMLGSDAKEDKCRKCGDGT 172  
QY 484 CRLVSGNLTDRGGPLGYQKILWIPAGALRLQIAQLRPPSSNYLALRGPGRSSTINGNWD 543  
DB 173 CKTIRNTITTKDLAPGYNDLLLPBEGATNIRIEEFPSSNYLACRNSHGHYLLNGDWRID 232  
QY 544 PPSYRAGTGVFRYRNPREECKGSLSAEGPTTQPDVYMIQFQENPGVYQYVYSSPP 603  
DB 233 FPRPMFFANSWNYORKPMGFAAPQLTCSGPISLSLIVLMVQEKNSLDYEYSI---- 288  
QY 604 PILENPTBPPVQLOPEILRVEPPLAPAPRPTFGTLQVRIPQMPAPPHPRTPLGS 663

Db 289 -----PE-----SLSHSQDTH----- 301  
QY 664 PAAYKRYGHSASCSGKGYVRPFLICISRESGEELDERSCAAGARPASPCHGTPC 723  
Db 302 -----THHQFNACSSCGGSGSRKVTNNRITLAENPCLDKSK-PVEEQACGTEPC 356  
QY 724 PYPWAGEWISCSRSCG-PGTQHRLOCRQFEGGGSSVPPE-----RCGLHPRNITQS 777  
Db 357 APHWVEGWSKSGCGSDGFGNQNSITCERISSGHEITVEDAVCLKEVGN--KPAKQE 414  
QY 778 C--QLRLCGHVEGSPWQCSVRCGRGORSQVRC-VGNNGDE--VSQECASGPPPPS 832  
Db 415 CNRDVKNCPKHYLG-PWTPCKLGDGQTKRKYTCFIEENGHKKRVLPEDCVC--EKPET 471  
QY 833 REACDMGPT-TAWPHSDWS 851  
Db 472 EKSCLLTCEGVWDWISQWS 491  
RESULT 6  
O44938  
ID O44938 PRELIMINARY; PRT; 1572 AA.  
AC O44938;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE THROMBOSPONDIN.  
GN THRI.  
OS Haemochus contortus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
OX NCBI\_TaxID=6289;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MOREDUN;  
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,  
RA Knox D.P.;  
RT "Cloning and characterization of Thrombospondin, a novel multidomain  
RT glycoprotein associated with the gut of Haemonchus contortus."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043121; AAB99830.2; -.  
DR HSP; P05067; ICA0.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00014; Kunitz\_BPTI; 12.  
DR Pfam; PF00090; tsp\_1; 6.  
DR PRINTS; PR00759; BASICPRASE.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00131; KU; 6.  
DR SMART; SM00209; TSP1; 7.  
DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 6.  
DR PROSITE; PS00092; TSP1; 2.  
KW Serine protease inhibitor.  
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2F903EC CRC64;

Query Match 14.2%; Score 695.5; DB 5; Length 1572;  
Best Local Similarity 28.3%; Pred No. 48-40;  
Matches 184; Conservative 70; Mismatches 230; Indels 167; Gaps 18;

QY 220 SVHTPSQAEPLSPETAQTEVAPRTAPLRHHHRAQASGTPEPSTHSLGEGGFRAFP 279  
Db 24 NVEPTLPHLPLSP-----PETNPAGAR----- 46  
QY 280 QPRPSSOGHASPQVAGRDPFPSPVPRGR-GQGGGPMGT-----GGTPHGPR 327  
Db 47 -AKRAQVQTV-----DGSVSLDKTGSGSGPWGVPWYEQCSRTCGGVQTEKR 96  
QY 328 LEPPDQHPGAWLPLLSNGPHASLWSLAFSPSPIPRCSEGEQLRACSAQPCPEQDPR 387  
Db 97 -----QCPG-----DCTGSPVRYVSCNLEPC-PEGTDFR 124

QY 388 ALQCAAFNSQEFMGQLYQYOMPEFTTEVQSGRCELNCRPRFRFRFYVRHTEKVDQTLCPGA 447  
Db 125 AEQCAAHNDPIDQYHKWIPY---KGKNCCELLCKPENGFYKWDOTVVDGKCDKSG 181  
QY 448 PDICVAGRLSPGCDGTLGSGRRPDCGCGGDDSCRLVSGNLTRDGGPLGQKILWIP 507  
Db 182 DDICVDGCVLPLGCDGLGSLALDKCGVCDGSGQCKTIEGSDERNLSPGHVWRIP 241  
QY 508 AGALRLQIAQLREFSSNYLALRPGGRSIIINGNAVDPGSGYRAGGVFRYRNRPRREGK 567  
Db 242 AGATAIRIEARPPSSNNLKMNSDYFLANGSMIOVEKDVNLGVHFEY-----DDGKP 296  
QY 568 ELSAEGFTTQPDVYVMIFQENPGVYQYVSSPPPILENPTPEPPVQLQPEILLRVEP 627  
Db 297 ERITAKGLEEEVVSVLIRKGRNDSIKYEFS----- 329  
QY 628 PLAPAPRPARTPGTLQQRVIRPOMPAPPHPTPLGSPAAYAKRWGH-SACASCSGKGVNR 686  
Db 330 -----VPIVDEVDMYKPGEWSACSVCCKGQQT 358  
QY 687 PIFLCISRESGEELDERSC-AAGARPPASPECHGTGTPCPYWEAGETWTSRSCG-PGTQ 744  
Db 359 RTPYCIETKQARVDDQLDDANSTKPEFEKPCETVDCQAEWFKGEWPCSCQTCGGQGEQ 418  
QY 745 HROLQCRQFEGGGSSVPPERCGLHPRPNTQSCQLRLCGHVEGSPWQCSVRCGRGQR 804  
Db 419 YRVYCHQVFANGRIITVDDGNCCTAERPVRQVRCNRFACPEWQAG-PWSACSEKCGDAFQ 477  
QY 805 SRQVRC-----VGNNGDEVSEQASGPPPPSREACDMGPT-TAWPHSDW 850  
Db 478 YRSVTSRSEKGERGKLLPACSS-ENTLESQRSCNLGPTGLKFFTEW 527  
RESULT 7  
Q96H81  
ID Q96H81 PRELIMINARY; PRT; 454 AA.  
AC Q96H81;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HYPOTHETICAL 51.1 KDA PROTEIN.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC00840; AAH0840.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 454 AA; 51054 MW; A674483B13A18F93 CRC64;

Query Match 13.5%; Score 562.5; DB 4; Length 454;  
Best Local Similarity 39.4%; Pred. No. 2.1e-38;  
Matches 132; Conservative 45; Mismatches 121; Indels 37; Gaps 6;

QY 554 VRYNRPR-BEGKGESLSABGPTTQPDVYVMIFQENPGVYQYVSSPPPILENPTPE 612  
Db 1 METYKRPNEISSTAGESFLAESPTNEILDVMIHQQPNPGVHYEYVINGTNAI----SPQ 56  
QY 613 PPVQLQP-----EILRVEPPLAPAPRPARTPGTLQQR 645  
Db 57 VP-PPRRPGEPPNGOMVTGKRSQEGQKNEEKEDLRGEAPEMFTSESQTPPVHRPD 115  
QY 646 VRIPQMP---APPHPRTPLGSPAAYAKRWGHSAASCSGKGVWRPFLICISRESGEELDE 702  
Db 116 RSPHPNPLVPAPQPRRSDHNNKQLGTETCSTCTGCKGQYPIFRVHRSHTHEAPE 175  
QY 703 RSCAAGARPPASPECHGTGTPCPYWEAGETWTSRSCSGPGTQHROLQCRQFEGGGSSVP 762

DB 176 SYCDSSMPTPEEPCNTPPCPAFWADIGENSECKTKGLGQHQHVLCRQVYANRSLTVQ 235  
QY 763 PERCHLPRPNITQSCRLRCLGHNVEGSPWQSVRCGRQSRQVRCVGNNGDEVSEQE 822  
DB 236 PYRCQHLEKPEPTTSCQLKICSEWQIRDTWTSVPCVGVGQRTDVKVCSNIGDWDVDEE 295  
QY 823 CASGPPQPPREACDMGCTTAFHSDWSSKVSPE 857  
DB 296 C-NMKLRPNNDIENCDMGCPAKSWFLTEWSERCSAE 329

RESULT 8  
Q9H824  
ID Q9H8E4 PRELIMINARY; PRT; 538 AA.  
AC Q9H8E4;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE PLACE2000373 PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y., Oshino A.;  
RT "NEDO human cDNA sequencing project.";  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK023772; BAB14673.1; -.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp.1; 6.  
DR SMART; SM00209; TSP1; 6.  
DR PROSITE; PS50092; TSP1.1.  
SQ SEQUENCE 538 AA; 60556 MW; 4C74B26D6FB75317 CRC64;

Query Match 13.5%; Score 661.5; DB 4; Length 538;  
Best Local Similarity 39.4%; Pred. No. 2.9e-38;  
Matches 132; Conservative 45; Mismatches 121; Indels 37; Gaps 6;

QY 554 VFRYNNRPPR-ERKGESLSAEGTTPQVDVYMFQENPVGVIYQVVISPPPILENTPPE 612  
DB 1 MFTYKRPNEISSAGESFLAEGTNEILDVYMHQPNPGVHYEYVINGTNAI-----SPQ 56  
QY 613 PPVPOLOP-----EILRVEPPLAPAPRPTPTGLQ 645  
DB 57 VP-PHRRPEPNQWVTGRSQEGRNEEKEDLGEAPEMFTSESACTPPVPHPD 115  
QY 646 VRIQMP-----APPHRPTPLGSPAAYKRVGHGSACSCKGKGVWRPFICISRESGEELDE 702  
DB 116 RFSHPHDLNVPAPQPPRRSRDHNNKQLGTTECTTCGKSGQYPIFRVHRSTHEAPE 175  
QY 703 RSCAAGARPASPCHGPPCPYPYEAQWTSCSRSCGPGTQHRLOCRQEGGGGSSVP 762  
DB 176 SYCDSSMKTPEEPCNTPPCPAFWADIGENSECKTKGLGQHQHVLCRQVYANRSLTVQ 235  
QY 763 PERCHLPRPNITQSCQLRCLGHNVEGSPWQSVRCGRQSRQVRCVGNNGDEVSEQE 822  
DB 236 PYRCQHLEKPEPTTSCQLKICSEWQIRDTWTSVPCVGVGQRTDVKVCSNIGDWDVDEE 295  
QY 823 CASGPPQPPREACDMGCTTAFHSDWSSKVSPE 857  
DB 296 C-NMKLRPNNDIENCDMGCPAKSWFLTEWSERCSAE 329

RESULT 9  
O76840

ID O76840 PRELIMINARY; PRT; 2167 AA.  
AC O76840; Q22911;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE C37C3.6 PROTEIN.  
GN C37C3.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;  
OC Rhabditiida; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Geisel C., Bradshaw H.;  
RT "The sequence of C. elegans cosmid C37C3.";  
RT Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -/- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY  
CC OF FORM B.  
CC -/- ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT  
DR EMBL; U64857; AAC25868.1; -.  
DR EMBL; U64857; AAC25867.1; -.  
DR HSP; P00981; 1DTK.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00047; ig; 1.  
DR Pfam; PF00014; Kunitz\_BPTI; 11.  
DR Pfam; PF00090; tsp.1; 6.  
DR PRINTS; PR00759; BASICTPASE.  
DR SMART; SM00408; IGC2; 1.  
DR SMART; SM00131; KSP; 11.  
DR SMART; SM00209; TSP1; 7.  
DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 10.  
DR PROSITE; PS0279; BPTI\_KUNITZ\_2; 11.  
DR PROSITE; PS50092; TSP1; 4.  
KW Alternative splicing; Hypothetical protein; Immunoglobulin domain;  
KW Serine protease inhibitor.  
FT VARSPPLIC 147 155 MISSING (IN ISOFORM A).  
FT VARSPPLIC 1556 1558 KDD -> SKF (IN ISOFORM A).  
FT VARSPPLIC 1559 2167 MISSING (IN ISOFORM A).  
SQ SEQUENCE 2167 AA; 237599 MW; 96274786D52E3639 CRC64;

Query Match 13.1%; Score 640.5; DB 5; Length 2167;  
Best Local Similarity 28.0%; Pred. No. 3.8e-36;  
Matches 149; Conservative 77; Mismatches 215; Indels 91; Gaps 12;

QY 344 NGPHASSLSLFAFSPPIPR-----CSGESEQLRACSQAPCPPEQDPDR 387  
DB 71 SGOKETGNWGPWVPENECSRSGGVQLEKRCGSDCTGASVRYISCLNAC-ESGTDPR 129  
QY 388 ALQCAAFNSQEFWQLYQWEPFTEVQSQRCCLNCRPRGFRFYVHTEKYQDGTLCQGA 447  
DB 130 AEQCKSKNDALDNGYHKWTPY---KGNKCELVKRPSGNFYFKWADRVVDGKCDUSKS 186  
QY 448 PDICVAGRCISPGCDGILGSGRRPDGCGVGGDDSTCLVSGNLDRGGPLGYQKILWIP 507  
DB 187 NDICVDGECPLVGCDGKLGSLKFDKCGKCDGSGTCKTIEGRFDERNLSPGVHDIKLP 246  
QY 508 AGALRLQALRPPSNVYALRGPGRSIINGNWAYDPGSGYRAGGVYFRYRNPRESGK 567  
DB 247 EGATNIQIEARKSTNLNALKNGSDHFYLNGLIQVEKEVEVGTIFVY-----DAEP 301  
QY 568 ELSAEGTTPQVDVYMFQENPVGVIYQVVISPPPILENTPPEPVPOLOQELRVEP 627  
DB 302 ETLQAQPLSELVALLFRKGRSDTAIKYFISIP-----LEEV----- 341  
QY 628 PLAPAPRPTPTGLQVRIQVRIQVRIQVRIQVRIQVRIQVRIQVRIQVRIQVRIQVRI 687  
DB 342 -----DYMFKFDNWTPCSVSCGKGQTR 364

```
QY 688 IFLCISRESGEBELERSCAA-GARPPASPECHCTPCPPYWEAGETWCSRSRG-PGTQH 745
  || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 365 NLYCIDGKNGKVEDDLCEENNAIKPEFEKSCVDCAEWFTGDWESCSCTDQDQY 424
QY 746 RQLOCRQEFGGGSGVPERGCHLPNITQSCQLRLCGHWEVSGPWSQSGVRGGRORS 805
  || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 425 RVVYCHQVFANGRRVTVEDGNCITVERPVKATCFACPEWQAG-PWSACEKCGDAFY 483
QY 806 ROVRC-----VGNNGDEVSEQACAGPPQP-PSREACDMGCPCTTAWF-HSDWS 851
  || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 484 RSVTCRSEKEGEGKLLAADACPADEQEKEDTERTCNLGPCEGLTFTVTEWN 535

RESULT 10
O60345 PRELIMINARY; PRT; 951 AA.
AC O60345;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0605.
GN KIAA0605.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN.
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Chara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL; AB011177; BAA25531.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSP1; 7.
DR PROSITE; PS00092; TSP1; 2.
SQ SEQUENCE 951 AA; 104678 MW; CC73753F232BC1A2 CRC64;

Query Match 12.6%; Score 618; DB 4; Length 951;
Best Local Similarity 25.2%; Pred. No. 5.8e-35;
Matches 192; Conservative 83; Mismatches 235; Indels 252; Gaps 28;

QY 319 TGGTPHGRLEPDPQHPGAWLPLLSNGPHASSL-----WSLFA----- 356
  || || || || || || || || || || || || || || || || || || || ||
Db 27 TGTSDNSPTNS-----LEGGDTAFWGWETKWTAFSRSCGGGVTQERHCL 75
QY 357 -----PSSPIP-----RCGSEQLRACQAPCPPEQDPDRALQCAAFNSQEFMGQYQMEP- 408
  || || || || || || || || || || || || || || || || || || || ||
Db 76 QQRKXSVPGPGNRTCTGTSKRYQLCRVQECPPDGRSFRFEEQCSFNSHVYNGRTHQMKPL 135
QY 409 -----FTEVQGSORCELNCP-RGFYFVYHTEKVDGTLQ-PPGAPDICVAGRLSPGCD 462
  || || || || || || || || || || || || || || || || || || || ||
Db 136 YPDYVHI--SSPCDLHCTTVQDQ--QLMPARDGTSCKLTLRQVGVSGKEPTGCD 191
QY 463 GILGSGRRPDGGCGVGGDDSTCLVSGNLTRDGGPLGYQKILWIPAGALRLQIAQLRPSS 522
  || || || || || || || || || || || || || || || || || || || ||
Db 192 GVLFSTHTLDCGICQDGGSSCTHTVGNRYKGNALHGYSLVTHIPAGARDIQIVERKKA 251
QY 523 NYLALRGPGRIINGNWNADPPGSRAGGTVFNRNPPREKGG-ESLSAEGPTQPDV 581
  || || || || || || || || || || || || || || || || || || || ||
Db 252 DVLALADEAGYFFNGYKVDSPKFNAGTVVYKRRPMDVYETGIEYVAGQPTNQGLN 311
QY 582 VYMIFOE--ENGVFQYVISSFPPILENPTPEPP-----VP-- 616
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 312 V-MVWNQNGKSPSIFETVLLQPP---HESRPQIYGFSESAESQGLDAGLAGFIHNP 367
QY 617 -----OLQPEILRVPEPLAPAP-----RPARPTGTLQR 644
```

```
Db 368 GSLYGQASSERLGLDNRFLGHPGLDMELQPSQGETNEVCEQAGGACGPPRGKGRDR 427
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 645 QV-----RIPQMPAPPHP 657
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 428 NVVTGTLTGDKDDEVDTHFASQEFTSANAIQDLLGAGSLDKFTLNTNETVNSIFAQAP 487
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 658 RTPLG----- 662
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 488 RSSLAESFFVDYEENGAGPYLLNGSYLELSDRVANSSEAPFPNVSTSLTSAANRTH 547
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 663 -----SPAAY--WKRVGHSACSACGKGVWRPFIPLCISRESGEEELDERSCAA 707
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 548 KARTRPKARKQGVSPADMYRWKLSHSEPCSATCTTGMSAYAMCV-RYDGEVDDSYCDA 606
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 708 GARPPASPECHCTPCPPYWEAGETWCSRSRGPGTQHRLQCRQEFGGG-GSSVPPERC 766
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 607 LTRPEPVHEFCAGRECQPRWETSWSSECSRTGEGYQFVYVRCWKMLSPGDFSSVYSLC 666
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 767 --CHLPRNITQSCQLRLCQ-HWEVGPSPWSQSVRCG-RQORSRQVRCVGNNGDEVSE 822
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 667 EAAEAVRPERKTCRNPACQPQWEM-SEWSECTAKGERSVVTDIRC-----SEDE 717
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 823 CASGP-PQPPSRACDMGPCTTAFHSDNSKSVSPPEPPAIS 863
  || || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 718 KLCDPNTRPVGKNTGPPCDRQWTVSDWG-----PCSGSC 753

RESULT 11
ID Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F25H8.3 PROTEIN.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7905398;
RA Wilton R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smauld N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprout J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSP; P15167; IDTH.
DR MEMOPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptidase.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 14.
DR SMART; SM00209; TSP1; 18.
```

DR PROSITE: PS50215; ADAM\_MEPRO; 1.  
DR PROSITE: PS50092; TSP1; 6.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN1.  
SQ SEQUENCE 2165 AA; 244337 MW; FCC3DA8AA9C4888 CRC64;

Query Match 12.6%; Score 618; DB 5; Length 2165;  
Best Local Similarity 26.5%; Pred. No. 1.4e-34;  
Matches 190; Conservative 90; Mismatches 254; Indels 184; Gaps 30;

QY 316 PWGTGTP-----HGP-RLEPD-PQHPGAMLPILNSGPHASSLW-----SLFA 356  
DB 571 PW-ADGTPCDESRSMECHGACVRLAPESLTKIDQWGDWRSWGECSRTCGGGVOKGLRD 629  
QY 357 PSSPIPR-----CSGESQLRACQAPCPPEQDPDRALQCAAFNSQEP-----MGOLYQW 406  
DB 630 CDSKPRNGKVCYQGERYRSCNTQCPWDTQPVREVCSEFNKIDIGIOGVASTNTHW 689  
QY 407 EP-FTEVGSQRCLENCPRGRFYVRTEKVDGTICQCAPDICVAGRCLSPGCGDIL 465  
DB 690 VKYANAVAPNERCKLYCLSGSAFYLLRDVKGDTGCDRNGDDICVAGACMPAGCDHQL 749  
QY 466 GGRPDGCGVCGGDDSCRLVSGNLTRDGGPLGYQKILWIPAGALRLQIAQ-----LRP 520  
DB 750 HSLTRDKCGVCGDDSCVKVKTENQ-GTFGYNEVWKIPAGSANIDIRKGYNNMKE 808  
QY 521 SNNYALRGPGGRSTINGNAVDPGYSYRA-GGVFYRNRPRREGKSGESLSAEGTTPQ 579  
DB 809 DDNYLSLAANGFEFLNGHFOVSLARQOIAFDQTVLEYS-----GSDAITERINGTCPIRS 865  
QY 580 VDVMYI-FOENPGVGYOYVLSRPPILENTPEPPVQLOPEILRVEPLAPAPRART 638  
DB 866 IYVHVLVSGSHPPDISYEMTAAPV---NAVIRPISALY--LWRVTDWTECDRACRG 919  
QY 639 P-----GTQLR 644  
DB 920 QOSQKWLCLDMSTRQSHDRNCQNVLPKPKQATRMGNIDCSTRWITEDVSSCAKCGSQK 979  
QY 645 OVRIP-----OMPAPPH-----PRTPLGSPAAY-----WKEVGHSAASAG-KGV 684  
DB 980 RRVSCVMKMGDRQTPASEHLCDNRNKSFDIASCIYDCSGRKNYGEWTSCTSETCGSNGK 1039  
QY 685 WRPIFLCSRESGEELDERSCAAGRAPSPAPCPCHGTGCPYWEAGETSRCSCGPGTQ 744  
DB 1040 MHRKSYCVD-DSNRVDESICGR-EQKATEBECNRIPC-PRWYVGHWSERSCSDGVK 1096  
QY 745 HRQLQCRQ-----EFG-----GGG-----758  
DB 1097 MRHAQDLAADRETHSRGCPAQTOEHCHNEHACTWQFVMSDCSAKCGDGVQYRDANCT 1156  
QY 759 -----SSVPPERCGLHPRNITQSCOLRLCGHWEVGSQVSCVRCGRQVRQVRCVGN 814  
DB 1157 DRHRSVLPHEHCLKMEK-IITKPCHEKSCPKYKLG-WSQVSCVEDGWSRRVSCVSGN 1214  
QY 815 GDEVSEECAGSPPOPPSREACDMGCPCTTANFHSWSSKVSPEPPATSCILGNHAQDT 872  
DB 1215 GTEVMSLCGTASDRPASHOTCNLGTG-CFWRNTDWSA-----CSVSCGIGHRETT 1265

RESULT 12  
Q56L37

AD Q96L37 PRELIMINARY; PRT: 1427 AA.  
IC Q96L37;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE VON WILLEBRAND FACTOR-CLEAVING PROTEASE PRECURSOR.  
GN ADAMTS13.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX PubMed=11557746;  
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,  
RA Fujikawa K.;  
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a  
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";  
RL J. Biol. Chem. 276:41059-41063(2001).  
DR EMBL; AV053376; AAL17652.1; --  
KW Signal; Protease.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.  
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABCA1442 CRC64;

Query Match 10.9%; Score 532.5; DB 4; Length 1427;  
Best Local Similarity 23.5%; Pred. No. 8e-29;

Matches 272; Conservative 86; Mismatches 430; Indels 371; Gaps 50;

QY 19 PQCLDDEVLSGHSLOTPTEEGQPGVWGWVQWASCSQPCGCVGQRRSRTQCLPTVOL 78  
DB 11 PPLCVAGILACGFL-----GCMGPFSHFOOCLQ-----ALEPQAVSSYL 50  
QY 79 HPSLPLPRPRHPEALLPRGQPRPTSPETLPLVYRQSRGRGGLR-----GPASH 131  
DB 51 SPQAPLGRPP-----SPGFQQRQRRRAAGGILHLELLVAVGPDVF 93  
QY 132 LGREETOEIRAAARR-----SRLRDPKPGMGYGRVFPFALPLHRNRHPRSPRSELSLI 186  
DB 94 QAHQEDTERVYVLTMLNIGAEALLRPSLGAQFRVHLVKWVI-LTEPEGAPNITANLTSLL 152  
QY 187 SSRGEEAIPPTPRAPFSAN-----GSPQTELPPTLSVHTSPQAEPLSPE-----234  
DB 153 SVCGWSOTIPEDDTPGHADLVLYTRFDELPDGNRQVRGVTVQLGGACSPWSCLITE 212  
QY 235 -----TAQTEV-----APTRPAPLRHHPRAQASGTEPP-----SPTH---267  
DB 213 DTGFDLGVTTAHEIGHSFGLHDGAPGSGCGPSGH---VMSDGAAPRAGLAWSPCSRQ 269  
QY 268 --SLGEGGFTRASPQPRRPSQGWASPOVAGRRDPDPSPVPRGQQGQ---GPWGTGT 322  
DB 270 LLSLLSAGRARCVMDPRP-----QFGSAGHPDPAQGLIYSANEQCRVAFGFKAVACT 323  
QY 323 PHGPRLE-----DPQHPGA-----WLPLL-----SNG-----PHAS-349  
DB 324 FAREHLDQALSCHTDPLDQSSCSLLVPLLDGTGCGVEKWCCKGRCSLVELTPIAAV 383  
QY 350 -SLWSLFAPSSPIPR-----CSGESEQLRACSQAPCPPEQP 384  
DB 384 HGRWSSWGRPSRSPSCGGGVVTRRRQNNPRPAFGRCVACVADLAQEMCTOACEKTOL 443  
QY 385 DPRALQCAAFNSQEFM-----GOLYQW-EPEFTEVQGSQRCLENCPRGRFYVRHTEKVQ 438  
DB 444 EFWSSQCARTDGQPLRSSPGGASFYHWGAAPHISQGDALCRHMCRAIGESFIMKRGDSFL 503  
QY 439 DGTLCQPGAP-----DICVAGRCLSPGCGDILGSGRRPDGCGVCGDDSTCLVSGNLT 492  
DB 504 DGTFCMPSGPREDTLSLVCVSGSCTFGCDGRMDSQOVMDRCQVCGGDNSTCSPRKGSFT 563  
QY 493 DRGGLGYQKILMIPAGALRLQIAQLRPSNNYLALRGPGGRSTINGNAVDPGYSYRA--550  
DB 564 -AGRAREYVTLVTPNLTSTVYIANHRLPETHLAVR-IGGRYVYVAKMISPTNTPSLL 621  
QY 551 --GGTVFY-----NRPPREGKSGESLSAEGTTPQVD--VYMIFQEE-----NPGVFYQY 597  
DB 622 EDGRVEYRVALTEDRLPRL-----EETIRNGPLQEDADIQVYRRYGEYGNLTPDITFY 677  
QY 598 VISSP-----PP-----604  
DB 678 FQPKPRQAWMAAVRGPCSVSCGAGLRWNYSLDQARKELVETVOCQSQQPPANPEAC 737  
QY 605 ILENPTP-----EPPV--PQLQPEILRVEPFLA-----630  
[1]

Db	738	VLE-PCPPYAVGVDFGSCASCGGGLRERPVRCVBAQSGLLKTLPPARCRAQAQPAVAL	796
Qy	631	-----PAPPAR-----TPGTLOQVRIP	649
Db	797	ETCNFQPCPARVEVSEPSCTSGAGLALENETCPVPGADGLEAPVTEGPGSVDEKLPAP	856
Qy	650	Q-----WPAPP-----HPPTPLGS-----PAAVWKRGVHGSACSACGKGVWRP	687
Db	857	EPCVGMSCPPGGHLDATSAEKAPSPGWSINTGAQAARVWTPVAGS-CSVSCGRGLMEL	915
Qy	688	IFLCISRESGEBLERSCAAGARPPASPEPCHTGTPPYWEAGETSCSRSCGPGTQHRQ	747
Db	916	RELCHDSALRPVPEELCGLASKPSRREVCOAVCPARWQY-KLAACSVSCGRGVVERI	974
Qy	748	LOCREFG--GGGSSVPPEPCGHLRPNITQSCURLC-GHWVGSPPWSQSVRCGRQOR	804
Db	975	LYCARAHGDDDEILLDTQCQGLRPERPEQAECSLEPCPRWKVMS-LGPCSASGLGTA	1033
Qy	805	SQVRVCG--NNGD--EVSBOECASGPPPPSREACDMGPTTAWPHSDSWSSKVSPEPAI	861
Db	1034	RSVACVQLDQDQVEYDEAACA-LVRPEASVPCLADCIYRHWVGTWME-----CSV	1086
Qy	862	SC-----ILGHAQ 870	
Db	1087	SCGDGIQRRRDCGLGPAQ 1105	
RESULT 13			
Qy	96R4	PRELIMINARY; PRT; 525 AA.	
AC	Q96R4		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	ADAM-TSRL RELATED PROTEIN 1.		
GN	ADAMTSRL		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	Hirohata S., Apte S.S.;		
RT	"A novel member of ADAM-TS related gene, ADAM-TSRL (A Disintegrin-like		
RT	And Metalloproteinase domain with Thrombospondin type I modules		
RT	Related gene-1).";		
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF176313; AAK84170.1; 3F180C3CAA7BA68 CRC64;		
SQ	SEQUENCE 525 AA; 58351 MW; 3F180C3CAA7BA68 CRC64;		
Query Match 10.8%; Score 530; DB 4; Length 525;			
Best Local Similarity 25.9%; Pred. No. 4e-29;			
Matches 149; Conservative 61; Mismatches 213; Indels 152; Gaps 18;			
Qy	351	LWSLAPSSPIPR-----CSGESEQLRACSAQPCPEQDPRALOCA 392	
Db	35	LWDAGPWSCESTCGGGAANSURCLSKSCGEGRNIRVTCNVNDCPPEAGDFRAQCS 94	
Qy	393	AFNSQFMQOLYQWEPFTEVQSGRQCELNCRPRGFRFYVHTEKVDGTLCPQGPADICV 452	
Db	95	AHNDVKHGGFYEWLPSN-DPDNPSCLCAQAGTTLVVELAPKVLDTGRCYTESLDMCI 153	
Qy	453	AGRLSPGCDGILGSRRRPDGCGVCGDSTCLVSGNTDR-GGPLGQKILWIPAGAL 511	
Db	154	SGLCQIVGCDHQLGTVKEDNCVCNGDGTCLRLVRQYKSLSATKSDTVAIPYSR 213	
Qy	512	RLQIQLRPSNYL---ALRGPGRGIINGNAVDPPGYSRAGGTVFYRNRPREEKGE 568	
Db	214	HIRLVKGPDLHYLEKTKLQTKGENSLST-----GTFLDVNSVDVQKPEPKD-----E 263	
Qy	569	SLSAEGTTPQVDVYVIFQENPG-----VFYQYVSSPPPLENTPPEPPVQLQP 620	
Db	167	AQGNLVVELAPKVLDTGRCNTDLSLDMISGICQAVGCDRLGSLNAKEDNCGVCGDST 226	
Db	264	ILRWAGPLTAD----FIVKIRNNGSADSTVQVIFYQ-----PIIHR-----	300
Qy	621	EILRVPEPLAPAPRPARTPGTLQQRIPOMAPPHPRTPLGSPAAWKRGVHGSACSASC 680	
Db	301	-----WRETDFFPCSATC 313	
Qy	681	KGWVRPIFLICISRESGEBLERSC---AAGARPPASPECHGTGTPC-----	724
Db	314	GGYQLTSAECYDLRSNRVADQCHYYPENIKPKLQECNLDPCPASDGYKQIMPYDL 373	
Qy	725	---PYWEAGETSCSRSCGPGTQHRQLC-ROBFGGGSSVPPPERCGHLRPNITQSCQ 779	
Db	374	YHPLPRWEATPWACSSCGGIGSRACVSEVEDIQGHVTSVEEMKCMYTPKMPIAQC 433	
Qy	780	LRLCGHVEVSPWSQSVRCGRQORQVRCVGNNGDEVSEQECASGPPPPPREACDM- 838	
Db	434	IFDCPKW-LAQEWSPCVTGGLRYRWVLCIDIRGHMHTG---CSPKTPHKECIIVP 489	
Qy	839	GPC-----TTAWFHS-----DWSKSVSPPEP 858	
Db	490	TPCYKPKLEKLPVEAKLPWFKQAQLESGAAVSEEP 524	
RESULT 14			
Qy	P82987	PRELIMINARY; PRT; 766 AA.	
AC	P82987		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)		
DE	ADAM-TSL3 PRECURSOR (FRAGMENT).		
GN	ADAMTSL3		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	Hirohata S., Anand-Apte B., Seldin M., Apte S.S.		
RT	"Punctin, a member of a new family with similarities to ADAM-TS		
RT	proteases, is a component of extracellular matrix of skeletal		
RT	muscle";		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
CC	-!- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.		
DR	EMBL; AF237652; AAK15041.1; "		
DR	InterPro; IPR000884; TSP1.		
DR	Pfam; PF00090; tsp.1; 6.		
DR	SMART; SM00209; TSP1; 7.		
DR	PROSITE; PS00092; TSP1; 1.		
KW	Repeat; Signal;		
FT	SIGNAL 1 26	POTENTIAL.	
FT	CHAIN 27 766	ADAM-TSL3.	
FT	DOMAIN 79 123	TSP TYPE-1 1.	
FT	DOMAIN 422 474	TSP TYPE-1 2.	
FT	DOMAIN 482 528	TSP TYPE-1 3.	
FT	DOMAIN 568 625	TSP TYPE-1 4.	
FT	DOMAIN 648 703	TSP TYPE-1 5.	
FT	DOMAIN 707 759	TSP TYPE-1 6.	
FT	NON_TER 766		
SQ	SEQUENCE 766 AA; 85987 MW; A42613E87AE91719 CRC64;		
Query Match 10.4%; Score 510.5; DB 4; Length 766;			
Best Local Similarity 25.7%; Pred. No. 1.4e-27;			
Matches 130; Conservative 57; Mismatches 188; Indels 131; Gaps 14;			
Qy	364	CSGESEQLRACSAQPCPEQDPRALQCAAFNSQEFMGQLYQWEPFTEVQSGRQCELNCR 423	
Db	108	CEGQNIKYKTCNSNHDPEADAEFRAQCSAYNDVQYQGHYYEWLPRYN-DPAAPCALKCH 166	
Qy	424	PGFFRYVHTEKVDGTLCPQGPADICVAGRCISPGCDGILGSRRRPDGCGVCGDST 483	
Db	167	AQGNLVVELAPKVLDTGRCNTDLSLDMISGICQAVGCDRLGSLNAKEDNCGVCGDST 226	

QY	484	CRVSGNLTDRGGP-LGYQKILWIPAGALRLQIAQLRPPSSNYL--- <td>539</td>	539
Db	227	CRVSGNLTDRGGP-LGYQKILWIPAGALRLQIAQLRPPSSNYL--- <td>539</td>	539
QY	540	WADPPGSGYRAGGVVFNKRNPPREGKESLSABGP-----TTPQVDYVIMFOE	588
Db	286	WADPPGSGYRAGGVVFNKRNPPREGKESLSABGP-----TTPQVDYVIMFOE	588
QY	589	ENPGVQYVIVSSPPPILENPTPEPPVQLOPEILRVEPPLAPAPRPTPGTLQROVRI	648
Db	335	ENPGVQYVIVSSPPPILENPTPEPPVQLOPEILRVEPPLAPAPRPTPGTLQROVRI	648
QY	649	POMAPPHPRTPGLSPAAYKRWGHSACSCGKGVWRPIFLCISRSGBELDERSC---	705
Db	340	POMAPPHPRTPGLSPAAYKRWGHSACSCGKGVWRPIFLCISRSGBELDERSC---	705
QY	706	AAGARPASPPCHGTCPC-----PYWEAGWTSRSCSGCGTQHQRL	748
Db	384	AAGARPASPPCHGTCPC-----PYWEAGWTSRSCSGCGTQHQRL	748
QY	749	QCROE-FGGGSSVPPRCGHLPRNITQSCQLRCLGHWEGVSPWQSCVRCGRGORSRQ	807
Db	444	QCROE-FGGGSSVPPRCGHLPRNITQSCQLRCLGHWEGVSPWQSCVRCGRGORSRQ	807
QY	808	VRCVGNNGDEVS-----EQEC 823	
Db	503	VRCVGNNGDEVS-----EQEC 823	

Search completed: July 23, 2002, 21:12:41  
Job time: 475 sec

RA	Stapleton R., Brownstein F., Young J., George R.,
RA	Chavez C., Dorsett V., Farfan D., Frise E.,
RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA	Nunoo J., Pacleab J., Paragas V., Park S., Phouanavong S., Wan K.,
RA	Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AY061825; AAL27636.1; "
SQ	SEQUENCE 1014 AA; 112241 MW; 030AI645935D5360 CRC64;
Query Match 10.3%; Score 505; DB 5; Length 1014;	
Best Local Similarity 22.6%; Pred. No. 4.5e-27;	
Matches 183; Conservative 86; Mismatches 229; Indels 310; Gaps 33;	
QY	307 RGRGQGGGPGWGTGGTGHGRLEPDQCHPGCAWLPLLNSNGPHASSLWSLF----- 355
DB	59 RSKVQSGAGGGGGAGG-----GPGQW-----SSWSDNSTCSTRCTDGGI 95
QY	356 -----APSSPIPRCGSGSEQLRACSQAPCPPEQDPDPRALQCAAFNSQEFMGQIYQWE 407
DB	96 MHQMRRCGSPGS---CRGESTRYRICNMQPC-PEQQDFRASSQCSAYNDVPYDGLTKWT 150
QY	408 PFTEVQGSQRCELNCRGRFRF-----YVRHTEKVQDGTLC 443
DB	151 PHYDY--VEPCALTCRGHPAHLVEDISRETGCDGNAEEAEHYDEQSVIVQLSARVQDGTFC 208



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 21:05:26 ; Search time 24.69 Seconds  
(without alignments)  
1375.336 Million cell updates/sec

Title: US-10-041-770-2

Perfect score: 4895

Sequence: 1 MENWGRPWLYLLLLSLPO.....PPAISCILGNHAQDTSAPFA 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query %	Length	DB ID	Description
1	709	14.5	1077	1	AT10_HUMAN	Q9h324 homo sapien
2	663.5	13.6	1593	1	AT12_HUMAN	P58397 homo sapien
3	584	11.9	1629	1	AT59_HUMAN	Q9p2n4 homo sapien
4	545	11.1	1211	1	AT52_HUMAN	Q95450 h adamts-2
5	538	11.0	450	1	AT10_MOUSE	P58459 mus musculu
6	531	10.8	1205	1	AT52_BOVIN	P79331 b adamts-2
7	514.5	10.5	860	1	AT56_HUMAN	Q9ukp5 homo sapien
8	503.5	10.3	967	1	AT51_RAT	Q9wuq1 rattus norv
9	501	10.2	1201	1	AT53_HUMAN	Q5072 homo sapien
10	486.5	9.9	968	1	AT51_MOUSE	P97857 mus musculu
11	466	9.5	967	1	AT51_HUMAN	Q9uh18 homo sapien
12	464.5	9.5	997	1	AT57_HUMAN	Q9ukp4 homo sapien
13	460	9.4	890	1	AT58_HUMAN	Q9up79 homo sapien
14	399.5	8.2	837	1	AT54_HUMAN	Q75173 homo sapien
15	399.5	8.2	905	1	AT58_MOUSE	P57110 mus musculu
16	395	8.1	930	1	AT55_HUMAN	Q9una0 homo sapien
17	385	7.9	930	1	AT55_MOUSE	Q9r001 mus musculu
18	373	7.6	630	1	AT54_RAT	Q9esp7 rattus norv
19	337.5	6.9	867	1	SSPO_BOVIN	P98167 bos taurus
20	277.5	5.7	2142	1	BAT2_HUMAN	P48634 homo sapien
21	262	5.4	660	1	YHL1_EBV	P03181 Epstein-bar
22	248	5.1	3530	1	M15_HUMAN	Q9ukn7 homo sapien
23	244	5.0	3511	1	M15_MOUSE	Q9qz24 mus musculu
24	243.5	5.0	1183	1	DRFL_RAT	P54258 rattus norv
25	238.5	4.9	963	1	YQ36_CAEEL	Q09457 caenorhabdi
26	235.5	4.8	620	1	EXTN_TOBAC	P13983 nicotiana t
27	235.5	4.8	1572	1	BA12_HUMAN	O60241 homo sapien
28	231	4.7	1736	1	CA2B_HUMAN	P13942 homo sapien
29	225.5	4.6	1584	1	BA11_HUMAN	O14514 homo sapien
30	224.5	4.6	1185	1	DRFL_HUMAN	P54259 homo sapien
31	216	4.4	1453	1	CALL_CHICK	P02457 gallus gall
32	214	4.4	2715	1	TRX2_HUMAN	Q9umnc6 homo sapien
33	212.5	4.3	1496	1	CAJ5_HUMAN	P05997 homo sapien

#### ALIGNMENTS

```

RESULT 1
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT: 1077 AA.
AC Q9H324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC similarity).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF163762; BAG35563.1;
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; zn_Mtpeptdse.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC Pfam: PF01421; Reprolysin; 1.
CC Pfam: PF00090; TSP1; 5.
CC SMART: SM00209; TSP1; 5.
CC PROSITE: PS00215; ADAM_WEPPO; 1.
CC PROSITE: PS00092; TSP1; 2.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix.
CC
CC NON_TER 1
CC PROPEP <1 207 BY SIMILARITY.
CC CHAIN 208 1077 ADAMTS-10.

```

P53420 homo sapien  
 P04258 bos taurus  
 P81122 mus musculu  
 P02452 homo sapien  
 P12105 gallus gall  
 Q98377 canis fami  
 P78559 homo sapien  
 P20508 homo sapien  
 P39060 homo sapien  
 Q9nzm4 homo sapien  
 P02458 homo sapien  
 P08121 mus musculu





FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	215	215	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1593 AA;	177545 MW;	07F9F48E63BD83A3 CRC64;	
Query Match 13.68; Score 663.5; DB 1; Length 1593;					
Best Local Similarity 29.88; Pred. No. 3.6e-23;					
Matches 168; Conservative 78; Mismatches 203; Indels 115; Gaps 20;					
QY	296	GRPPFPFVPRGRQQGGPWG-----TGGTPHGPRL--EPDPHPGAWLPLLSNGPH	347		
DB	535	GKKPE---SIPGGWGR--WSPWHSRSCGAGVQSAERLCHNNPPKFGKY-----	580		
QY	348	ASSLSLAPSPPIPCGSESEQLRACSOACPPEPPRALQCAAFNSQEFMGLQIWE	407		
DB	581	-----CTGKRKRYLCLNVPCHSEAPTFROMQSEFTVPYKNELYHWF	624		
QY	408	PTEVQSGORCELNCRGFRGVYRHTKVDGTLCPQG--APDVCVAGRLSPGCCDIL	465		
DB	625	PI--FNPAPHCLELYCRPIDGQFSEKMLDAVIDGTFCFEGGSRNVCIKMGWCDVEI	682		
QY	466	GSRRPDGCGVGGDDTCRLVSGNLTDRGGPLGYQKILNIPAGALRLOIAQLRPSSNYL	525		
DB	683	DSNATEDRGVCLGDSGSCQTVRKFKOKES--GYVDIGLIPKPGARDIRVMEIEGAGNFL	741		
QY	526	ALRGPG--GSIINGNWAYDPPGSRAGGTFRYRNPREEKGESLSAEGTTPQVDVYM	584		
DB	742	AINSEDPKXYIINGFIQWNGNYKLAGTVQYDR---KGLDKLMATGPTNESVWLQL	797		
QY	585	IFQENPNGVFYQVISSPPPILENTPPEPPVQLQPEILRVEPLAPAPRPTGTLQR	644		
DB	798	LQVTPNGIKYETIKDQ--LDND-----	820		
QY	645	QVRIQMPAPPHPRLPLGSPAAYKRVGH--SACSASCGKGVWRPFLCISRESGELDER	703		
DB	821	---VEQM-----YFQYGHWTCSVTCGTGIRQTAHCICKRG--MYKAT	861		
QY	704	SCAAGARPASPEPGCHGTGTPPYWEAGWTSRSCGP--GTQHQLOCRQEFGGGSSVP	762		
DB	862	FCDPETQPNRGKCKEKACPPRWAGEWACATCGPHGEKKRVLCITQTMVSDQALP	921		
QY	763	PERCHLPNPNTQSCQLR--LC--GHWEVGSFWSQSVRCGRGQRORQVRCVGNNGDEVSE	820		
DB	922	PTDCQHLKPKTLSCNRDILCPDWTGVN--WSECSVSGGGVRSVTCAKNH-----	974		
QY	821	QECASGPPQPPSREACDMGPCTTA	844		
DB	975	DEPCDVTFRPNSEALCGLGQQCFSS	998		
RESULT 3					
AT99_HUMAN	STANDARD;	PRT;	1629 AA.		
ID	AT99_HUMAN				
AC	Q9P2N4; Q9NR29;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase				
DE	with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).				
GN	ADAMTS9 OR KIAA1312.				
OS	Homo sapiens (Human).				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. [1] NCBI\_TaxID=9606; SEQUENCE FROM N.A. (SHORT ISOFORM). RC TISSUE=Fetal; MEDLINE=20396138; PubMed=10936055; RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A., Maki R.A.; RT \*ADAMTS 9, a novel member of the ADAM-TS/Metallospodin gene family. RT Genomics 67:343-350(2000). [2] SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM). RC TISSUE=Brain; MEDLINE=20181126; PubMed=10718198; RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.; RT \*Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro. RT DNA Res. 7:65-73(2000). CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY). CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY). CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING. CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES. EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG, PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR THYMUS. CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY). CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY). CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B. CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN. CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS. ----- this SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)). ----- EMBL; AF261318; AAF89106.1; -; EMBL; AB037733; BAA92550.1; -; MIN; 605421; -; InterPro; IPR001762; Disintegrin. InterPro; IPR002870; Pep\_M12B\_propep. InterPro; IPR001590; Repolysin. InterPro; IPR000884; TSP1. InterPro; IPR000130; zn\_Mtpeptdse. Pfam; PF01562; Pep\_M12B\_propep; 1. Pfam; PF01421; Repolysin; 1. Pfam; PF00090; tsp\_1; 11. SMART; SMO0209; TSP1; 12. PROSITE; PS02015; ADAM\_MEPRO; 1. PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG. PROSITE; PS50092; TSP1; 9. PROSITE; PS00142; ZINC\_PROTEASE; 1. Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix; Alternative splicing. SIGNAL 1 18 POTENTIAL. PROPEP 19 287 BY SIMILARITY. CHAIN 288 1629 ADAMTS-9. DOMAIN 509 587 DISINTEGRIN-LIKE. DOMAIN 589 642 TSP TYPE-1 1. DOMAIN 645 752 CYS-RICH. DOMAIN 753 880 SPACER.

DOMAIN	999	1053	TSP TYPE-1 2.	
DOMAIN	1056	1108	TSP TYPE-1 3.	
DOMAIN	1111	1156	TSP TYPE-1 4.	
DOMAIN	1184	1239	TSP TYPE-1 5.	
DOMAIN	1240	1295	TSP TYPE-1 6.	
DOMAIN	1332	1383	TSP TYPE-1 7.	
DOMAIN	1386	1439	TSP TYPE-1 8.	
DOMAIN	1445	1498	TSP TYPE-1 9.	
DOMAIN	1501	1554	TSP TYPE-1 10.	
DOMAIN	1562	1612	TSP TYPE-1 11.	
DOMAIN	88	96	POLY-SER.	
SITE	223	223	CYS-STEINE SWITCH (POTENTIAL).	
METAL	434	434	ZINC (CATALYTIC) (BY SIMILARITY).	
ACT_SITE	435	435	BY SIMILARITY.	
METAL	438	438	ZINC (CATALYTIC) (BY SIMILARITY).	
METAL	444	444	ZINC (CATALYTIC) (BY SIMILARITY).	
CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).	
CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL).	
CARBOHYD	271	271	N-LINKED (GLCNAC. . .) (POTENTIAL).	
CARBOHYD	749	749	N-LINKED (GLCNAC. . .) (POTENTIAL).	
CARBOHYD	840	840	N-LINKED (GLCNAC. . .) (POTENTIAL).	
CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).	
CARBOHYD	1267	1267	N-LINKED (GLCNAC. . .) (POTENTIAL).	
VARSPLIC	1064	1072	CLVTGKGH -> VRWEGCYFP (IN SHORT ISOFORM).	
VARSPLIC	1073	1629	MISSING (IN SHORT ISOFORM).	
CONFLICT	367	367	F -> L (IN REF. 1).	
SEQUENCE	1629 AA; 182649 MW;	CIC4CEFF58B8941F CRC64;		
Query Match 11.9%; Score 584; DB 1; Length 1629;				
Best Local Similarity 26.2%; Pred. No. 1.3e-19;				
Matches 191; Conservative 73; Mismatches 235; Indels 230; Gaps 31;				
QY	316	PW--GTGTP-----HG----	PRLEPDPOHGAWLPLLSNGPHASSLSLFPASS-----	359
DB	559	PNADGTECEPKHCKYGCVPKEMDVPVTDGWS-----	GSWSPFGTCSRTCGG	607
QY	360	-----PIPR-----	CGSEQLRCAOAPCPPPODPALQCAAFNSQEF--MGQ	402
DB	608	IKTATRECNREPKNKGKYGVRMRKFKSCNTECLKQKDRFDEQCAHDFGKHFNGL	667	
QY	403	L--YQWEP-FTEVQSGQRCNLNCRPRGFRFVNRHTEKVQDGTLCQGPADICVAGRCCLSP	459	
DB	668	LPNVRWPKYSGILMKDRCKLFCRVAGNTAYQLRDVRIDGTGCGDNTDVCVQGLCQRA	727	
QY	460	GDGILGSRPDDGGVCGGDDSTCRVLSGNLTDRGGPLGVOKILWIPAGALRLQIAQLR	519	
DB	728	GCDHVLSKARRDKCGVCGGDSNCKTVAG--TFNTVHYGNTVVRIPAGATNIDVRQHS	785	
QY	520	PS-----SNYLALRPGGRSIINGWAYD-PPQSYRAGGTFRYNRPPEEKGKGSLSAE	573	
DB	786	FSGETDDNLYALSSSGKEFLNGFVVTMAKREIRIGNAVVEYS--GSETAVERINST	842	
QY	574	GPTTOP--VDVYMFQENPGVYQYVLISSPPILENTPPEPPVQLOPETLLRVEPLAP	631	
DB	843	DRIEQELLQVLVSGKLYNPVRSFNI-----PIEDKP-----	QQFYWNSHGFWQA	889
QY	632	APRPARTPTGLQRY-----	RIPQMAPPHRPTPLGSPA-AYWKRVGHSAC	676
DB	890	CSKPCQ--GERKRLVCTRESQTLTVSDQCDRLPQPGHITTEPCGTDCDLRHHVARSSEC	947	
QY	677	SASCKGWVRPFLC--ISRESG--EELDSRCAAGARPAPSEPCGHT-----	721	
DB	948	SAQCGLGRTYLDIYCAKYSRLDGKTEKYDDGFCSSHPK-PSNRKCSGCNTGGWRSYSAW	1006	
QY	722	-----	PCPPYWEAGETWCS	736
DB	1007	TBCSKSCDGGTORRAICVNTNRNDVLDKSKTHQEKVTIQRCSFPC-PQWKGSDWSECL	1065	
QY	737	RSCGPGTOHRLQCRQETG-----		
DB	1066	VTCGKGHHRQWVC--QFGEDRLNDRMCDPDKPTSMQTCQOPECASWQAGFWGQCSVTC	1123	

QY	756	GGG-----SSVPPERCGLPRPNITQSCQLRLC-----	783	
DB	1124	GQGYQLRAVKCIIGTYKSVYDDNDNCAATPTDQCELPSCHPPAAPETRRSTYSAPR	1183	
QY	784	GHWVGSPPSQSVRCGRGORSQVRCVGNNGDEVSEQCASGPPOPPPREACDMGCPCTT	843	
DB	1184	TQWRFGS-WTPCSATCGKTRMYVSCRDNGSVADESACAT-LPRPVAKESCVTPC-G	1240	
QY	844	AWPHSDWS 852		
DB	1241	OKWALDWS 1249		
RESULT 4				
ID	ATS2_HUMAN	STANDARD;	PRT: 1211 AA.	
AC	O95450;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)			
DE	(Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI)			
DE	(Procollagen I/II amino-propeptide processing enzyme).			
GN	ADAMTS2 OR PCINP OR PCPNI.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).			
RC	TISSUE=Skin;			
RX	MEDLINE:99347935; PubMed:10417273;			
RA	Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E., Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W., Byers P.H., Lapierre C.M., Prockop D.J., Nusgens B.V.; "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene."; Am. J. Hum. Genet. 65:308-317(1999).			
CC	-!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN COLLAGEN BIOSYNTHESIS.			
CC	-!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-1-Gln.			
CC	-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	-!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).			
CC	-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI; ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-PROCOLLAGEN PEPTIDASE ACTIVITY.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.			
CC	-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	-!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).			
CC	-!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS SYNDROME TYPE VIIC (EDS-VIIC), A RECESSIVELY INHERITED DISORDER CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.			
CC	-!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.			
CC	-!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AJ003125; CAA05880.1; -  
 DR MIM; 604539; -  
 DR MIM; 225410; -  
 DR InterPro; IPR001762; Disintegrin.  
 DR InterPro; IPR002870; Pep\_M12B\_propep.  
 DR InterPro; IPR001590; Reprolysin.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR000130; Zn\_Mtpeptidse.  
 DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
 DR Pfam; PF01421; Reprolysin; 1.  
 DR Pfam; PF00090; tsp; 1; 4.  
 DR SMART; SM00209; TSP1; 4.  
 DR PROSITE; PS00215; ADAM\_MEPRO; 1.  
 DR PROSITE; PS00092; TSP1; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;  
 KW Alternative splicing.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 253 BY SIMILARITY.  
 FT CHAIN 254 1211 ADAMTS-2.  
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 409 409 BY SIMILARITY.  
 FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DOMAIN 480 560 DISINTEGRIN-LIKE.  
 FT DOMAIN 561 617 TSP TYPE-1 1.  
 FT DOMAIN 618 722 CYS-RICH.  
 FT SITE 691 693 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DOMAIN 723 851 SPACER.  
 FT DOMAIN 852 911 TSP TYPE-1 2.  
 FT DOMAIN 912 974 TSP TYPE-1 3.  
 FT DOMAIN 975 1030 TSP TYPE-1 4.  
 FT DOMAIN 40 43 POLY-ALA.  
 FT DOMAIN 185 188 POLY-GLU.  
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 544 566 HCFKHCILTPDILKRDGWSGA -> FRPGAVALHACYPS  
 FT VARSPLIC 567 1211 TLGGGGRWIA (IN ISOFORM SPNPI).  
 FT SEQUENCE 1211 AA; 134722 MW; BECEEF25C23CAD2D CRC64;

Query Match 11.1%; Score 545; DB 1; Length 1211;  
 Best Local Similarity 26.9%; Pred. No. 5.8e-18;  
 Matches 167; Conservative 71; Mismatches 247; Indels 136; Gaps 23;

QY 289 WASPVAGRRPDPFSPVPRGRGQGGMGTTGTPHGPRLPDPQHPGAWPLLLS----- 343  
 DB 552 WLTPDILK-----DGSW-----GAWSPFGSRTC 577  
 QY 344 -----NGPHASSLWLFAPSPIPRCGSEQLRACSAQPCPEQDPRALQCA 392  
 DB 578 GTGVKTRTCQDNPH-----PANGGRTCSGLAYDFQLASRQCDPSLADFREQCR 628  
 QY 393 AFNSQ--EFWGQLYQWPFTEVQSGRCCLNCRPRGRFRVVRTEKVTQGTLCQ--PGAPDI 450  
 DB 629 QMDLYFEHGDAQHHLWLPHEHRAKAKCHLYCSRETGEVVMKRMVHDGTCRSYKDAFSL 688

QY 451 CVAGRCISPGCDGILSGRRPDDGCGVGGDDSTCRVLSGNLTDGGLGQKILWIPAGA 510  
 DB 689 CVRGDCRKVGCDGVGSGKQBDKCGVGGDNCKVYKGTFTSPKKHGYIKMFEIPAGA 748  
 QY 511 LRLQIAQLRPSNYSIALRG-PGGRSIIINGNWAYDPPG-SYRAGCTVFRNRPPEEGKGE 568  
 DB 749 RHLLIQEVDATSHLAVKNLETKFILNEENDVDASSKTFIANGVEWEY-----RDEGRE 804  
 QY 569 SLSAEGTTQPDVYMI-FQENPGVYQYVSISSPPFIENPTPEPPVQLOPEILRVEP 627  
 DB 805 TLQTMGLPHGTIVLVPVGDTRVSLTYKMI-----HEDSLNVD- 844  
 QY 628 PLAPAPRPPARTPGTLQQRVIRPQMPAPPHPRTPLGSPAAYKVRVGHSAACSASCGKGVWRP 687  
 DB 845 -----DNNVLEED-----SVVEWALKKWKSPCKPCGGGSGQFT 877  
 QY 688 IFLCISRESGEBELDERSCAAGAPASPECHGTPC-PPYWEAGEWTSCRSQG-PQTQH 745  
 DB 878 KYGRRRLDHKKVHRGFCALSKPKAIRACNPQECSPVWVTGWEWPCQTCGTGMQV 937  
 QY 746 ROLQROEFGGGS-SVPPPERGCHLPRNITQSCQLRLC-GHWEVGSWQSCSVRCGRGQ 803  
 DB 938 RSVKCIQPLHNTTRSVHAKCND-ARPESSRACRELCPGRWRAG-PWSQCSVTCNGT 995  
 QY 804 RSRQVRC-VGNNGDEVSEQECAGPPQPPPSREACDMGPC-----TAMFHSWSSKVS 855  
 DB 996 QERPVPCKRTADDSFGICQEE-----RPETARTCLGCPCRNISDPSKKSYYVQWLSRPD 1049  
 QY 856 PEPPAISCILGNHAQ-DTSAP 875  
 DB 1050 PDSPIRKISSGHGCGQDKSIF 1070

RESULT 5

ID AT10\_MOUSE STANDARD; PRT; 450 AA.  
 AC PS8459;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-10 (EC 3.4.24.-) (A disintegrin and metalloproteinase with  
 DE thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).  
 GN ADAMTS10.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Seldin M.F., Apté S.S.;  
 RT "A novel gene of the ADAMTS family predicts ADAMTS-10, a  
 RT metalloprotease with unique structural features and expression  
 RT pattern";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -1- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF302012; AAK97226.1; -

DR PROSITE; PS00092; TSPL; 1.  
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Repeat;  
FT Extracellular matrix.  
FT NON\_TER 1 51  
FT DOMAIN 52 174  
FT DOMAIN 171 232  
FT DOMAIN 234 290  
FT DOMAIN 294 348  
FT DOMAIN 353 399  
FT CARBOHYD 86 86  
FT CARBOHYD 141 141  
FT CARBOHYD 238 238  
SQ SEQUENCE 450 AA; 48861 MW; 2377DUE2CFBFFCA CRC64;  
  
Query Match 11.0%; Score 538; DB 1; Length 450;  
Best Local Similarity 32.0%; Pred. No. 5.3e-18;  
Matches 141; Conservative 52; Mismatches 192; Indels 56; Gaps 15;  
  
QY 437 VODGTLCPGAPDTCVAGRCISPGCDGILSGRRPDGCGVGGDDTCRLVSGNLTDRGG 496  
DB 4 VVDGTPCPDPTVDCVSGECHKVHGVGDRVGLSDREKCRVCGDGSACETIEGVFSPALP 63  
QY 497 PLUGYOKILWIPAGALRLQIAOLRPSNLYALRGPGGRSINGNNAVDPPGSRAGGTVER 556  
DB 64 GTGYEDVWIPKSGVHFIQDLNLSLHALKGQESLLEGLPTGPQPHRLPLAGTTFH 123  
QY 557 YNRPREGKSGESLAEGTPTQVDVYMFQENPGVYQVVISPP-----PPILNPT 610  
DB 124 LRQGP---DQAQSLGALGPINSLTIWYLAQALPALHYRF--NAPIARDALPPYSWHYA 178  
QY 611 PEPVPVQ-----LOPEILRVEPLP---APAPRPARTGTL-ORQVRIPQMPAPPHPRTPL 661  
DB 179 PWTKSAQACGSGVQVYVCRNQLDSSAVPHYCSGSKLPKRGACNTECPDP----- 233  
QY 662 GSPAAYKRVGHSACSGKGVWKPFIPLCISRESGEE---LDRSCAAGARPASPPEPC 718  
DB 234 -----WVGNWNRSCSDAGVRSRVVQCRRVSAEKAALDSACQP-PRPPVL-EAC 285  
QY 719 HCTPCPPYWEAGFWSCRSRCPGQTHQLOCRQEFGGSGSVPERGHLPRNITQSC 778  
DB 286 QQPMCPPEWATLWSECTPSCGGLHRRVVLCKS--ADQRTLPFGHCLPAKPPSTWRC 343  
QY 779 QLRLC--GHEWYSGSWSCVRCGRQSRQVRCVGNNGDEVSEQECASGPPQPPSREAC 836  
DB 344 NLRRCPPARW-VTSEWGECSQGLGQQQRTVRCSTHTGQ-----PSRECT 388  
QY 837 D-MGPCTTANFHSWSSKVSP 856  
DB 389 EALRPSTMQQCEAKCDSVYPP 409  
  
RESULT 6  
AT52\_BOVIN  
ID AT52\_BOVIN STANDARD; PRT; 1205 AA.  
AC P79331;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and  
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)  
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I  
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).  
GN ADAMTS2 OR NPI.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP PROPEP 29 253  
RC TISSUE=Skin; ADAMTS-2.

RX MEDLINE-97225960; PubMed-9122202;  
RA Collige A., Li S.W., Sieron A.L., Nussgens B.V., Prockop D.J.,  
RA Lapierre C.M.;  
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a  
RT new member of the superfamily of zinc-metalloproteinases with binding  
RT sites for cells and other matrix components.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).  
RN [2]  
RP PARTIAL SEQUENCE.  
RX MEDLINE-95348096; PubMed-7622483;  
RA Collige A., Beschip A., Samyn B., Goebels Y., Van Beeumen J.,  
RA Nussgens B.V., Lapierre C.M.;  
RT "Characterization and partial amino acid sequencing of a 107-kDa  
RT procollagen I N-proteinase purified by affinity chromatography on  
RT immobilized type XIV collagen.";  
RL J. Biol. Chem. 270:16724-16730(1995).  
CC -I- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR  
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO  
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN  
CC COLLAGEN BIOSYNTHESIS.  
CC -I- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
CC alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains  
CC at Ala-|-Gln.  
CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -I- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO  
CC COLLAGEN TYPE XIV.  
CC -I- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -I- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL  
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS  
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS  
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.  
CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -I- PTM: THE N-TERMINUS IS BLOCKED.  
CC -I- PTM: THE PRECURSOR IS CLEAVED BY A FUFIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -I- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOPARAXIS, A  
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN  
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN  
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -I- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
CC -I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X96389; CAA65253.1;  
DR InterPro: IPR001762; Disintegrin.  
DR InterPro: IPR002870; Pep\_M12B\_propep.  
DR InterPro: IPR001590; Reprolysin.  
DR InterPro: IPR000884; TSPL.  
DR InterPro: IPR000130; zn\_Mtpeptdse.  
DR Pfam: PF01562; Pep\_M12B\_propep; 1.  
DR Pfam: PF01421; Reprolysin; 1.  
DR Pfam: PF00090; tsp.1; 4.  
DR SMART: SM00209; TSPL; 4.  
DR PROSITE: PS00215; ADAM\_MPEPRO; 1.  
DR PROSITE: PS00092; TSPL; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 28  
FT PROPEP 29 253  
FT CHAIN 254 1205  
FT ADAMTS-2.

FT	METAL	402	402	ZINC (CATALYTIC) (POTENTIAL).
FT	ACT_SITE	403	403	BY SIMILARITY.
FT	METAL	406	406	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	412	412	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	474	551	DISINTEGRIN-LIKE.
FT	DOMAIN	555	611	TSP TYPE-1 1.
FT	DOMAIN	612	716	CYS-RICH.
FT	DOMAIN	717	845	SPACER.
FT	DOMAIN	846	905	TSP TYPE-1 2.
FT	DOMAIN	906	968	TSP TYPE-1 3.
FT	DOMAIN	969	1024	TSP TYPE-1 4.
FT	SITE	685	687	CELL ATTACHMENT SITE (POTENTIAL).
FT	DOMAIN	31	35	POLY-ALA.
FT	DOMAIN	177	180	POLY-GLU.
FT	CARBOHYD	104	104	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	245	245	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	942	942	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	943	943	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	987	987	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1025	1025	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1092	1092	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1139	1139	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1144	1144	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1205 AA;	133887 MW;	7B5B232A45320371 CRC64;
Query Match 10.8%; Score 531; DB 1; Length 1205;				
Best Local Similarity 26.2%; Pred No. 2,4e-17;				
Matches 164; Conservative 70; Mismatches 235; Indels 156; Gaps 22;				
QY	289	WASPVAGRRPDP	PPSVPRGRCQGQGGPMTGGTGHGPRKLEPDHPHGWLPLLS-----	343
Db	546	WLTPDLKR-----	-----DGNW-----	-----GAWSPFGSCSRTC 571
QY	344	-----NGPHASSLWSL	FAPSPICRSGESELACRACSOAPCPPEQDPDRALQCA 392	
Db	572	GTGVKPTRCDNPH-----	-----PANGRTCSGLAYDFOLCNSQDCPALADFREEQCR 622	
QY	393	AFNSQ-EFMQQLQW	PEFFTEVQSRCENCRPRGRFRFYRHTKQVDTLQCO-PGAPDI 450	
Db	623	QWDLFEHGDAQH	HLWHPHEHRDAKERCHLYCESKETGEVVMKRMVHDTRGSKYKDAFSL 682	
QY	451	CVAGRCLSPCD	GILSGRRPDCGVCGGDDSTCLVSLGNLDRGGPLGCKLWIPAGA 510	
Db	683	CVRGDCRKVC	GDVIGSSKOEDKCGVCGGDNHCKVVKGTFSRSPKLLGIYKMFPA 742	
QY	511	LRLIQALRPS	NKALRG-PGGRSIINGNVAVDPPG-SYRAGTGYFRYNRPREEGKE 568	
Db	743	RHLIQEADT	TSHLAVKNLETKFILNEENDVDPSKTFIANGVEWEX-----RDEGDE 798	
QY	569	SLSAEGTTPQ	VDVYIMFQ-EENPGVFYQVVISPP-----PILENPTPEPPVPLQPE 621	
Db	799	TLQTMGPLHG	TIIVLVIPEGDARISLYKYMIEDSLNVDNNVLED-----	845
QY	622	ILRVEPLAPAP	PTPGTLQQRVIRPQMPAPPHPRTPLGSPAAYWKRVGHSACSASCG 691	
Db	846	-----	-----DSVGEWALKWSPCKPCG 865	
QY	682	KGYWRPFLC	ISRESGELDERSCAAGRPASPCHGTPC-PPYWEAGETWSCRSRG 740	
Db	866	GGSQFTYGR	RRLDHKWHVRFCDVSFKPKAIRTCNPQECSPQVYVWTEWEPSCRSRG 925	
QY	741	-PGTQHQRLQ	CRQEFGGGS-SYPPERCGLHPRENTQSCQLRLC-GHWEEVGSFWSOCV 797	
Db	926	RTGMQVRSV	RCVQPLHNHTFRSVTHKIND-ARPEGRACNRELCPGRWRAGS-WSQCSV 983	
QY	798	RCGQRQSRQ	VRQC-VGNNGDEVSEQACSGAPPPPSREACDMGCP-----TTAFHSD 849	
Db	984	TCGNGTQEP	RVLCRTADDSFGVCREE-----RPTARICRLGCPRNTSDPSKKSYYVQ 1037	
QY	850	WSSKVSVEPP	PAISILGNHAQDTSA 874	
Db	1038	WLSRPDPNSP	-----VOETSS 1053	



FT DOMAIN 796 852 TSP TYPE-1 2.  
 FT DOMAIN 68 71 POLY-ARG.  
 FT DOMAIN 662 665 POLY-GLY.  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 843 843 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 850 AA; 97098 MW; E57213015DEC265 CRC64;

Query Match 10.58; Score 514.5; DB 1; Length 860;  
 Best Local Similarity 31.28; Pred. No. 1e-16;  
 Matches 125; Conservative 44; Mismatches 147; Indels 85; Gaps 9;

QY 304 SVPRGRGQGGPGTGTGTPHGRLEDDPQHPGAWLPGLLSNGPHASSIW----- 352  
 DB 475 SIPAAEGTLQ-----TGNIEKGWCYQDCVFGTWPQSIDGWCWPSLWGCSTCGGV 530  
 QY 353 --SLFAPSPIPR-----CSGESQLRACSOAPCPPEPPDPRALQCAAFNSQEFMGQLYQ 405  
 DB 531 SSSLRHCDSPAPSGGKYCLGERKRYRSCNTDPCPLGSRDFRERQCADFDMNPRGRKYN 590  
 QY 406 WEPFTEVGSORCELCNPRGFRFVRHTEKVDGTLCOPGAPDICVAGRCLSPGCCDIL 465  
 DB 591 WPIYTG-GGVKPCALNCAEINGFYTERAPAVIDGTQCNADSLDICEGCKHVCNDIL 649  
 QY 466 GSGRRPDGCGVGGDDSTCRVLSNLTDP---RGGLPGYQKILWIPAGALRQIAQLRPS 522  
 DB 650 GSDAREDCRCVGGGGSTCDATIEGFENDSLPRG---GYMEVVQIPRGSHVIEVREAVMSK 706  
 QY 523 NYLALRGPGGRIINGWAVDPGYSRAGGVFRVNRPPREEGKSGESLSAGPTQPDV 582  
 DB 707 NYIALKSGDDYYINGAWTIDWPRKFDVAGTAFYKRPTEP---ESLEALGPTSENIV 763  
 QY 583 YMIFQENPGVYQVVISPPPILENTPPEPPVQLOPEILRVEPPLAPAPRPTARTQTL 642  
 DB 764 MVLLQEQNLGIRYKENV-----PITRTGSG 788  
 QY 643 QRQVRIPQMPAPPHPRTPLGSPAAYWKRVGHSACSASCGKG 683  
 DB 789 DNEVGFT-----WNHPWSECSATCAGG 811

RESULT 8  
 ATSL\_RAT  
 ID ATSL\_RAT STANDARD; PRT; 967 AA.  
 AC Q9W001; Q9ER11;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).  
 GN ADAMTS1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;  
 RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,  
 RA Little S.P.;  
 RA "Induction of a disintegrin and metalloproteinase with the  
 RT thrombospondin type I motif (ADAMTS).";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 18-967 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=20304099; PubMed=10847486;  
 RA Diamantis I., Luethi M., Hoessli M., Reichen J.;  
 RT "Cloning of the rat ADAMTS-1 gene and its down regulation in

endothelial cells in cirrhotic rats.";  
 RL Liver 20:165-172(2000).  
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE  
 CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY  
 CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH  
 CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER  
 CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-|-LEU-1684  
 CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.  
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM  
 CC CIRRHOTIC LIVER.  
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF149118; AAD34012.1; -;  
 CC EMBL: AF304446; AAG29823.1; -;  
 CC InterPro: IPR001762; Disintegrin.  
 CC InterPro: IPR002870; Pep\_M12B\_propep.  
 CC InterPro: IPR001590; Repolysin.  
 CC InterPro: IPR000884; TSP1.  
 CC InterPro: IPR000130; Zn\_MTPeptidse.  
 CC Pfam: PF01562; Pep\_M12B\_propep; 1.  
 CC Pfam: PF01421; Repolysin; 1.  
 CC Pfam: PF00090; tsp\_1; 3.  
 CC SMART: SM00209; TSP1; 3.  
 CC PROSITE: PS0215; ADAM\_MEPRO; 1.  
 CC PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 CC PROSITE: PS00092; TSP1; 2.  
 CC PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 CC Repeat; Extracellular matrix; Heparin-binding.  
 CC SIGNAL 1 54  
 CC PROPEP 55 252  
 CC CHAIN 253 967  
 CC SITE 205 205  
 CC METAL 401 401  
 CC ACT\_SITE 402 402  
 CC METAL 405 405  
 CC METAL 411 411  
 CC DOMAIN 476 558  
 CC DOMAIN 559 615  
 CC DOMAIN 616 724  
 CC DOMAIN 725 857  
 CC DOMAIN 858 907  
 CC DOMAIN 908 967  
 CC DOMAIN 194 198  
 CC CARBOHYD 547 547  
 CC CARBOHYD 720 720  
 CC CARBOHYD 764 764  
 CC CARBOHYD 782 782  
 CC CARBOHYD 945 945  
 CC CONFLICT 21 21  
 CC CONFLICT 26 31  
 CC CONFLICT 49 49  
 CC CONFLICT 72 72  
 CC CONFLICT 79 79

BY SIMILARITY.  
 ADAMTS-1  
 CYSTEINE SWITCH (POTENTIAL).  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 BY SIMILARITY.  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 DISINTEGRIN-LIKE.  
 TSP TYPE-1 1.  
 CYS-RICH.  
 SPACER.  
 TSP TYPE-1 2.  
 TSP TYPE-1 3.  
 POLY-ARG.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 I -> V (IN REF. 2).  
 KFRSQ -> RSRGSL (IN REF. 2).  
 V -> A (IN REF. 2).  
 R -> P (IN REF. 2).  
 L -> TR (IN REF. 2).





FT	METAL	404	404	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	466	546	DISINTEGRIN-LIKE.
FT	DOMAIN	547	603	TSP TYPE-1 1.
FT	DOMAIN	604	708	CYS-RICH.
FT	DOMAIN	709	840	SPACER.
FT	DOMAIN	841	898	TSP TYPE-1 2.
FT	DOMAIN	899	961	TSP TYPE-1 3.
FT	DOMAIN	962	1013	TSP TYPE-1 4.
FT	DOMAIN	242	245	POLY-ARG.
FT	CARBOHYD	79	79	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	115	115	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	238	238	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	341	341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	471	471	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	810	810	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	938	938	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1201 AA;	135113 MW;	D54EA92BD506A3AA CRC64;
SQ	SEQUENCE	1201 AA;	135113 MW;	D54EA92BD506A3AA CRC64;
Query Match 10.2%; Score 501; DB 1; Length 1201;				
Best Local Similarity 27.0%; Pred. No. 5, 3e-16;				
Matches 187; Conservative 68; Mismatches 272; Indels 166; Gaps 31;				
QY	227	QAEPSPETAQTEV-APRTAPAPLRHRAQASGTEPPSPTHS----	LGEGGFFRASPOP	281
Db	407	QGNRCGETANGVMAFLVQAAPHRYH-WSRSGQELKRYIHSYDCLDD-----		455
QY	282	RRPSSQGWASQVAGRRPDPFPSPRGSGQGGPWGTC-----		320
Db	456	--PFDHWP-----KLPE-LPGINTSMEQCFDFGVGYKMTAFRFDPCQKQWCHSP		506
QY	321	-----OTPGHLEPDPQHPGAWL-----	PLLSNPHASSLWLSFAPS----	359
Db	507	DNPFYCKTKKGLDTECAAGKWCYKGHCMWKNANQKQDGNWS--WTKFGSCSRTC		564
QY	360	-----PTP-----RCGSEQLRACSAQCPPEQDPALQCAAFNSO-EPWG		401
Db	565	TGVFRTRQCNPNPMPINGGDCFGVNFVQLCNTBECQKHFEFRAQCCQQRNSHFETQN		624
QY	402	QLYQWEPTEVQSGRCELKNCRRP--GFREYVHTKVKVQDGLQPGAP-DICVAGRCL		458
Db	625	TKHMLPYEHPDPKCHLYCSQKSGDAYMK--QLVHDGTHCSYKDPYSICVYRGECVK		682
QY	459	PGCDGILGSRRRDGCVCVGGDSTCLVSGNLTDRGGLGYQKILWIPAGALRLQIAQL		518
Db	683	VGCDKEIGSNKVEDKGVCGDGNHCHRTVKTFTPTPKLGLKMFDPDPPGARHVLQED		742
QY	519	RPSSNYLALRGP-GRSIINGNAVDPGSGYRAGGTVFYRNPPREKGSLSAEGPTT		577
Db	743	EASPHILAIKQATGHVILNGKEEAESRTFFIDLGVEDYN-----IEDDIESLHTDGLH		798
QY	578	QPVDMYIFQEN--PGVYQYVI--SSPPPILENPTPEPPVLPQLPTELLRVEPPLAP		633
Db	799	DPVILVLIQENDTRSLTYKIIHEDSVPTINSNV-----IQEELDTFE-----		844
QY	634	RPARTPOTLRQVRIPQMPAPPRTPLGSPAAYWKRVRGHSACSACGKGVWRPFLCIS		693
Db	845	-----WALKSWSQVSRKPCGGGFOYTKYGC-R		869
QY	694	RESGEELDEERS-CAAGARPPASPEPCHTGCP-PYWEAEWTSKCSRSCG-PQTHRQLQC		750
Db	870	RKSNKWHVSFCANKPKPIRMKNIQECTHPLWVAEWEHCTKCGSSGYQLTYVC		929
QY	751	QRE-FGGGGSPVPERGHLPRNITQSCQLRLC-GHNEVGSFWSOCVRCGRGSRQV		808
Db	930	LQPLDGTNRSVHSKYCMG-DRPESRRPCNRVPCPAQWKTG-PWSECSVTGEGTGVQV		987
QY	809	RCVGNNGDSEVSEQBCASGPPQPPSRACDMGPC		841
Db	988	LC--RAGDHCDGE-----KPESVRACQLPPC		1011

RESULT 10

ATSL_MOUSE	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
ID	P97857	054768;		
AC	30-MAY-2000	(Rel. 39, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DE	ADAMTS-1 precursor (EC 3.4.24.-)	(A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS1) (ADAM-TS1).		
GN	ADAMTS1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVJ;			
RX	MEDLINE=98110583; PubMed=9441751;			
RA	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RT	"The exon/intron organization and chromosomal mapping of the mouse			
RL	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97150761; PubMed=895297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.;			
RA	Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-			
RT	disintegrin family protein with thrombospondin motifs as an			
RT	inflammation associated gene.";			
RL	J. Biol. Chem. 272:556-562(1997).			
RN	[3]			
RN	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RX	MEDLINE=99303657; PubMed=10379500;			
RA	Kuno K., Terashima Y., Matsushima K.;			
RT	"ADAMTS-1 is an active metalloproteinase associated with the			
RT	extracellular matrix.";			
RL	J. Biol. Chem. 274:18821-18826(1999).			
RN	[4]			
RP	FUNCTION.			
RX	MEDLINE=20389568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.;			
RA	Ohno H., Matsushima K.;			
RT	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RL	FEBS Lett. 478:241-245(2000).			
RN	[5]			
RP	FUNCTION, AND INDUCTION.			
RX	MEDLINE=2043757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.;			
RA	Richards J.S.;			
RT	"Progesterone-regulated genes in the ovulation process: ADAMTS-1 and			
RL	cathepsin L proteases.";			
CC	Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).			
CC	INVOLVED IN ITS TURNOVER, HAS ANGIOGENIC INHIBITOR ACTIVITY (BY			
CC	SIMILARITY), ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH			
CC	VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER			
CC	CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY			
CC	SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 1691-GLU- -LEU-1692			
CC	SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.			
CC	-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR			
CC	MATRIX.			
CC	-1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY			
CC	INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY			
CC	LIPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA			
CC	CELLS OF PREOVULATORY FOLLICLES.			
CC	-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT			
CC	FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	-1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.			
CC	-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.			



RX MEDLINE-20181126; PubMed-10718198;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.,  
RT "Prediction of the coding sequences of unidentified human genes. XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:65-73(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20389799; PubMed-10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordsiek G., Horstschler K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dargatz E.,  
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehrach H., Reinhardt R., Zasp M.-L.; 21.";  
RT "The DNA sequence of human chromosome 21.";  
RL Nature 405:311-319(2000).  
RN [6]  
RP SEQUENCE OF 418-967 FROM N.A.  
RC TISSUE-Melanoma;  
RA Blum H., Bauersachs S., Meves H.-W., Weil B., Wiemann S.;  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CLEAVES AGGREGAN, A CATIONIC PROTEOLYCAN, AND MAY BE  
CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR  
CC ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH  
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER  
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.  
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-|-LEU-1939  
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AF170084; AAF15317.1;  
DR EMBL; AF060152; RAD48080.1; ALT\_INIT.  
DR EMBL; AF207664; RAF23772.1;  
DR EMBL; AB037767; BAA92584.1; ALT\_INIT.  
DR EMBL; AP001697; BAA95502.1;  
DR EMBL; AL162080; CAB82413.1;  
DR MIM; 605174;  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR002870; Pep\_M12B\_propep.  
DR InterPro; IPR001590; Repolysin.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
DR Pfam; PF01421; Repolysin; 1.  
DR Pfam; PF00090; tsp.1; 3.  
DR SMART; SM00209; TSP1; 3.  
DR PROSITE; PS0215; ADAM\_MPRO; 1. FALSE\_NEG.  
DR PROSITE; PS00427; DISINTEGRIN\_1;  
DR PROSITE; PS50092; TSP1; 2.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 49  
FT PROPEP 50 252  
FT CHAIN 253 967  
FT CHAIN 198 198  
FT SITE 401 401  
FT METAL 402 402  
FT ACT\_SITE 403 403  
FT METAL 405 405  
FT METAL 411 411  
FT DOMAIN 476 559  
FT DOMAIN 560 616  
FT DOMAIN 617 724  
FT DOMAIN 725 849  
FT DOMAIN 850 908  
FT DOMAIN 909 967  
FT DOMAIN 843 846  
FT CARBOHYD 547 547  
FT CARBOHYD 720 720  
FT CARBOHYD 764 764  
FT CONFLICT 227 227  
FT CONFLICT 468 468  
FT CONFLICT 561 561  
SQ SEQUENCE 967 AA; 105383 MW; C189389324741ED1 CRC64;  
  
Query Match 9.58; Score 466; DB 1; Length 967;  
Best Local Similarity 22.98; Pred. No. 1.6e-14;  
Matches 136; Conservative 71; Mismatches 179; Indels 208; Gaps 22;  
  
QY 318 GTGG-----TPHGPRLEPDQHPGAWLPLLSNGP-----HASSLSLWLFAP 357  
DB 513 GTSGLVLCQKHPWADGTSCGEGK--CINGKCNKTKDKHFDTPFHGS--WGMWGP 567  
QY 358 -----SSPIPR-----CSGESEQLRACSOAPCEPQDP--RALQCA 392  
DB 568 WDCSRCTGGGVQVYTMRECDNPKNGKGYCEGRKRVYRNCLEDCPDNNGKTFRESC 627  
QY 393 AFNSQEF-----MGQLYQWEP-FTEVQGSQRCNELNPRGRFYVRYHTEKVDGTLCP 445  
DB 628 AHN--EFSKASFGSGPAVEWIPKYAGVSPKDRCLICQAKGIGYFFVLQPKVVDGTPCSP 685  
QY 446 GAPDICVAGRLSPGCDGILGSGRRPDGCGVGGDDSTCRVLSGLNLTDRGGPLGYQKILW 505  
DB 686 DSTSVYGVQCVCACDRIDSKKKFKDCGCGGNGSTCKKISGVT--SAKPGYHDIIT 743  
QY 506 IPAGALRLQIAQ-----LRPSSNYLALRPGGRSINGNW---AVDPGSGYRAGTGVFRY 557  
DB 744 IPTGATNIEVKQRQGRSNGNSFLAKAAGTYILNGDYTLSTLEQDINXK--GVVLYR 801  
QY 558 NRPPRECKGESLSAE-----GPTQTP--VDVYMFQENPGVYQYVVISPPPILENPT 610  
DB 802 S-----GSSAALERIRSFSLKEPLTIQVLTGVALRPKITYTFVK-- 844  
QY 611 PEPVPVQLQPEILRVEPLAPAPRPPARTPGTLQQRVQPMPAPPHPRTPLGSPAAYWKR 670  
DB 845 ----- 844  
QY 671 VGHSAASACGKGVWRPIFLCISRESGEELDERSCAAGARPPASPECHGTPCPYWEAG 730  
DB 845 -----KESFNAIPTFSAWVIE 861  
QY 731 EWTSCSRSCGPGTQHQLOCRQETGGGSSVPPBCGHLPRPNITQSCQLRGLCGHWVGS 790  
DB 862 EWGCKSKCELGWRRLVECDRNG-----QPASECAKEVKASTRPCADIPCFQWQLGE 916  
QY 791 FWSQCSVRCGRQSRQVRQVGNNGDEVSEQECASGPPQPPFSR--EACDMGSPCT 842  
DB 917 -WSSCKTKCGKGYKRSRLKCLSHDGGVLSHESC--DPLKKPKHFDICTMAECS 967  
  
RESULT 12

AT57\_HUMAN  
ID AT57\_HUMAN STANDARD; PRT; 997 AA.  
AC Q9UKP4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).  
GN ADAMTS7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP  
RX MEDLINE=9939124; PubMed=10464288;  
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;  
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of  
Zinc Metalloproteases".  
RL J. Biol. Chem. 274:25555-25563(1999).  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,  
LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF140675; AAD56358.1; -.  
CC HSP; P15167; IATL.  
CC MIM; 605009; -.  
CC InterPro; IPR001762; Disintegrin.  
CC InterPro; IPR002870; Pep\_M12B\_propep.  
CC InterPro; IPR001590; Reprolysin.  
CC InterPro; IPR000384; TSP1.  
CC InterPro; IPR000130; Zn\_Mrpeptdse.  
CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
CC Pfam; PF00421; Reprolysin; 1.  
CC Pfam; PF00090; tsp-1; 1.  
CC SMART; SM00209; TSP1; 2.  
CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC PROSITE; PS0215; ADAM\_MEPRO; 1.  
CC PROSITE; PS50092; TSP1; 1.  
CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix.  
FT SIGNAL 1 27 POTENTIAL.  
FT PROPEP 28 232 BY SIMILARITY.  
FT CHAIN 233 997 ADAMTS-7.  
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).  
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 389 389 BY SIMILARITY.  
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DOMAIN 462 537 DISINTEGRIN-LIKE.  
FT DOMAIN 538 594 TSP TYPE-1 1.  
FT DOMAIN 595 697 CYS-RICH.  
FT DOMAIN 698 914 SPACER.  
FT DOMAIN 915 990 TSP TYPE-1 2.  
FT CARBOHYD 94 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;  
  
Query Match 9.5%; Score 464.5; DB 1; Length 997;  
Best Local Similarity 22.6%; Pred. No. 2e-14;  
Matches 217; Conservative 92; Mismatches 295; Indels 355; Gaps 44;  
  
QY 86 PRPPRH-----PEALLPRGQGRPQTS-----PETLPLRYTQSRGGRGLRGA 129  
DB 172 PAREGHAQPHVYKRAQAPERLAQRGDSSAPSTCGVQVYPE-----LESR----- 215  
QY 130 SHLGRETEQIRAAARSRLRDPKPGMGYGRVFPALPLHNRNRHPSRPPR----- 180  
DB 216 -----REWEQRQQRPRRLR-----RLHORSVSKWKWETLVV 249  
QY 181 SELSLISSREEAIPSPTRAEPE-----SANGSP-----QTEVPTRPAP---LRH- 251  
DB 250 ADAMVEYHGQ-----PQVESVLTMMVAGLFHDPDSIGNPIHITIVRLVLEDEE 302  
QY 216 PTELSVTTPSP-----QAEPLSPETA-----QTEVAPTRPAP---LRH- 251  
DB 303 DLKITHHADNTLKSFKWKQKSIINMKGDAPLHDDTALLTRKDLCAAMNRPCTGLSHV 362  
QY 252 ----HPRQAASGTEPP-----SPTHSLG-----EGFEFRASPOPRRPSSQGWASPV 294  
DB 363 AGMCQPHRSCSINEDTGLPLAFTVAHELHSGFIQHDGSGNDCEPVGKRPFIM---SPQL 419  
QY 295 -----AGRR-----PDP-----PFSVPRG-----RQOQG 313  
DB 420 LYDAAPLTWSRCSQYITRFLDRGWLCLDDPPAKDIIDFPSPVPPGVLYDVSHQCLQYG 479  
QY 314 -----QGPNGTGTGHPRLPDQ-----PGAWLPL-----LSN 344  
DB 480 AYSAFCEMDNVCHTLKCSVGTTCCHSKLDAVDGTRGKKNKLSGECVPVGRPEAVDG 539  
QY 345 GPASSLSLFLAPS-----SPIPR-----CSGESEOLRACSAQAPCPPEPDP 386  
DB 540 GWSGSAWSICSRSGMGVSAERQCTQTPKYKGRYCVGERKFRCLNQLQACAPGRPSF 599  
QY 387 RALOCFAFNSQEFMGOLYONPEPTEVQSGRCELNCRPRGFREYVVRHTEKVDGTLQCP- 445  
DB 600 RHVQCSHFDAKLYKGQHTWVTVV-----VNDVNPCELHCRPANEYFAKKLDAVDGTPQYV 657  
QY 446 -GAPDICVAGRCLSPGCGDILGSGRRPDGCGVCGDDSTCLVSGNLTDRGPGVGYKIL 504  
DB 658 RASRDLCLNGICKNVGCDFEIDSGAMEDRCVCHGNGSTCHTVSGTPEEAEG-LGYVDVG 716  
QY 505 WIPAGALRLQIAQLRPSNYLALRPG-GRSILINGNAVDPPGSRAGGTVFRNRPRE 563  
DB 717 LIPAGAREIRIQEVAEAAFLALRSEDPKTYFLNGGWTQWNGDIQVAGTFTTAR---- 772  
QY 564 EGKGESLSAEGTTPQVDVYVIFQENPQVGYQYVVISPPPILENTPPEPPVLPOLQPEIL 623  
DB 773 RGNWENLTSPGTTKEPYWI-QVPASRPGGSGRGGVPRPSTLHGRSRPGVSGSV----- 827  
QY 624 RVEPPLAPAPPARTPTGLQVQVRIPOMAPPHPTPLGLSPAAYKWKVGHSAACSACGK 683  
DB 828 -TEPGSEPGP-PAAASTSVSPSLKWPNLVAHV-----RGWG--QAPLGLG 870  
QY 684 VWRPFLICISRESGEELDERSCAAGARPPASPECHGTPCPPYWEAGETWISCSRCGPG- 742  
DB 871 CWRRLVLM-----GPRLPQ-----LLFQESNPGV 896  
QY 743 ----TQHRLQCRQFEFGGSGSVFPERGHLPRNITQSCOLRLCGHWEVGSPPSQSVR 798  
DB 897 HYEVTIHRE-----AGGHDEVP-----PVFS-----WHY3-PWTKTIVT 930  
QY 799 CGRQGR-----SRQVRCVGNNGDEVSEQCASGPPQPPPSREADM 838  
DB 931 CGRGEKWRGHSPTCRGLVSGGQHLQLPAHCWATLTGLEVFSE-----PQSICEM 981

RESULT 13  
AT58\_HUMAN  
ID AT58\_HUMAN STANDARD; PRT; 890 AA.  
AC Q9UP79; Q9NZS0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DT ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)  
DE (METH-8).  
GN ADAMTS8 OR METH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI-TaxID-9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=99367466; PubMed=10438512;  
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,  
Lombardo M., Iruela-Arispe M.L.;  
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new  
RT family of proteins with angio-inhibitory activity.";  
RL J. Biol. Chem. 274:23349-23357(1999).  
[2]  
RN SEQUENCE OF 195-440 FROM N.A.  
RP MEDLINE=20079168; PubMed=10610729;  
RX Georgiadis K.E., Hirohata S., Seidlin M.F., Apte S.S.;  
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on  
RT mouse chromosome 9 and human chromosome 11.";  
RL Genomics 62:312-315(1999).  
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER  
CC KIDNEY.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
CC  
CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL; AF060153; AAF48081.1; -;  
CC EMBL; AF125283; AAF25806.1; -;  
CC HSP; P34179; 1IAG.  
CC  
CC MIM; 605175; -;  
CC InterPro; IPR001762; Disintegrin.  
CC InterPro; IPR002870; Pep\_M12B\_propep.  
CC InterPro; IPR001590; Repolysin.  
CC InterPro; IPR000884; TSP1.  
CC InterPro; IPR000130; Zn\_MTPeptidse.  
CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
CC Pfam; PF01421; Repolysin; 1.  
CC Pfam; PF00090; tsp\_1; 2.  
CC SMART; SM00209; TSP1; 2.  
CC PROSITE; PS0215; ADAM\_MEPRO; 1.  
CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC PROSITE; PS00092; TSP1; 1.  
CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 27  
FT PROPEP 28 214 BY SIMILARITY.  
FT CHAIN 215 890  
FT ADAMTS-8.  
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 365 365 BY SIMILARITY.  
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DOMAIN 439 526 DISINTEGRIN-LIKE.  
FT DOMAIN 527 583 TSP TYPE-1 1.  
FT DOMAIN 584 690 CYS-RICH.  
FT DOMAIN 691 832 SPACER.  
FT DOMAIN 833 890 TSP TYPE-1 2.  
FT DOMAIN 202 205 POLY-PRO.  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 195 195 E -> R (IN REF. 2).  
FT CONFLICT 413 440 YLTSLDGGHGGCLLDADGAALPLTGL -> FSGHQLGW  
FT IFHKYLKCVSLKCDLMP (IN REF. 2).  
SQ SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;

Query Match 9.4%; Score 460; DB 1; Length 890;  
Best Local Similarity 24.9%; Pred. No. 2.8e-14;  
Matches 169; Conservative 76; Mismatches 225; Indels 208; Gaps 28;

QY 150 DPTKP-----GMGCGYGVPEALPLHRRNRHPRSPR-----SEL-----SLISSRGEA 193  
DB 375 DSKPCRTLFPGMGKHVMAPLFVHLNQTLPWSPCSAMYTELLDGGHGGCLLDADG-AA 433  
QY 194 IPSPTPRAEPFSANGSPQTELPPELTVHPTSPQAPL-----SPETAQTEVAPTRP 246  
DB 434 LPLPTGL-----FORMALYQDQOCROQIFGPFDRHCPNTSAQDVC----- 473  
QY 247 APLRHHRAQAGTEPSPHSLGEGFFFRASQPRPS---SQGWASPOVAGRRDPDP 303  
DB 474 AQLMCH-----TDGAEPCLHTKN---GSLWADGTGCGHLCSGSCLEPVEVERKP-- 524  
QY 304 SVPRGRQQGQGWGT-----GGT---PHGPRLEPDPQHPFGAWLPLLSNGPHASSLSLF 355  
DB 525 -VVDG-GWAPGWGECGRTRCGGVQVFSHRECKDPEPONGRY----- 565  
QY 356 APSPIPRCSGESEQLRACSAQCPPEQPPRALQCAAFNSQEFM---GOLYQWEP-FYE 411  
DB 566 -----CLGRRAKYQSCHTCECPDGKSFREQQCEKNAYNTMDGNLQWPKYAG 617  
QY 412 VQGSQRCCLNCRPGRFYVRHTEKVQDGLTQCPGADPCVAGRCISPGCDGILGSGRRP 471  
DB 618 VSPDRCKLFCRAGRSEFKVFEAKVIDTGLCGPETLAICVRGCQVKAGCDHVVDSPKRL 677  
QY 472 DGCVCVGGDDSTCLVSGNLTDRGGPLGYQKILWIPAGALRLQIAQ-----LRPSSNYLA 526  
DB 678 DKCVCGGKGNKSCRVKSGSLTPTN--YGYNDIVTIPAGATNIDVQRSHFPGVQDGNLYA 735  
QY 527 LRPGPGRSILNGWAVDP-FGSYRAGGTVFYRNPREEKGESLSAEGTTPQVDYMI 585  
DB 736 LKTDADGQILLNGLAISAEODILVKITLY-----SGSIATLERLSQF-- 780  
QY 586 FQENPGVFYQYVISSPPPILENPTPEPPVQLOPEILRVEPPLAPAPRPTGTLQOR 645  
DB 781 -----RPLPLETLVQL----- 791  
QY 546 VRIPQMPAPHPPTPLGSPRAYWKRVCCHSACSACSGKGVWRPIFLCI-SRESBELDERSC 705  
DB 792 LTVPEGVFPP-----KVKIT-----FFVNDVDFSMQSK---ERAT 825  
QY 706 AAGARPPASPEPCHGTCPPPYWEAGETSCSRCPGQTQHRQLQCRQFEGGGSSVPPER 765  
DB 826 TNIQLPLLHAQ-----WVLGDWNSCCSGAGWQRRTVECRDPSQASAT----- 870

```

QY 766 CGHLPRPNITQSCQLRLC 783
DB 871 CNKALPEDAKPCQSLC 888

RESULT 14
ATSA4_HUMAN
ID ATSA4_HUMAN STANDARD; PRT; 837 AA.
AC 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADAMS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMS4 OR KIA00688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RA "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteases.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=1082174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
CC site.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -1- INDUCTION: BY INTERLEUKIN-1.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12b.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

-1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN
-1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AB014588; BAA31663.1; -
EMBL; AF148213; AAD41494.1; -
EMBL; AY044847; AAL02262.1; -
HSSP; P34179; 1IAG.
MEROPS; M12.221; -
MLM; G03876; -
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_Mtpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS02032; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 362 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT DOMAIN 247 252
FT CARBOHYD 68 68
FT CONFLICT 77 77
FT CONFLICT 626 626
FT CONFLICT 682 682
SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;

Query Match 8.28; Score 399.5; DB 1; Length 837;
Best Local Similarity 25.38; Pred. No. 1.4e-11;
Matches 168; Conservative 70; Mismatches 264; Indels 161; Gaps 35;

QY 61 GVGQRRSRTCOLPTVQL--HPSLPPLPPRRHPHALLPR-----GQPR-PQTSF---E 109
DB 235 GAGLKYLLTMAAAAKAFKHPSI-----RNPVSLVWTRLVILGSGEGPQVGPSSAAQ 287
QY 110 TLPLYRTQSRGGLRGCPASHLG-----RETFQETRAARRSLRDPKRGMEGYGRVP 163
DB 288 TLRSFCAMQRLNTPEDSDPDHFTAILFTQDLGCVSTC-----DTL--GMADVGV- 338
QY 164 FALPLHRRNRPRSPRSELSISRG--EEAIPSPTRAEPFSA---NGSPQTEL-PPTE 218
DB 339 -----CDPARSCAIVEDDGLQSAFTAHELGHVFNLMHDSKPCISLNGPLS 385
QY 219 LSVHTPPSQAPLPSPETAQTEVAPTRPAPLRRHPRAQASG-----TEPPSPFH----- 267
DB 386 TSRHYMAPVMAHVDPPEWS-----PCSAFTITFDLNGYGHCLLDKPEAPLHLPTVF 438
QY 268 -----SLGEGGFRRASPPRRPSSQGWASPOVAGRRPDPFFSVPRGRQQGQGP 316
DB 439 FGDYDADRCQQLIFGPDSPRHCPLPPCAALWCSGHLNG-----HAMCQTRKSP 488

```



QY 317 WTGTGTPHGR-----LEPD-----PQHGAWLPLLSNGPHASSLSWLPAPS-- 359  
DB 489 W-ADTGPCGAACMGRCLEHWDQIQDNIPQ-AGGW-----GPGWDCSRT 535  
QY 360 -----PIPR-----CSGSEQLRACSQAPCPPEQP-DPRALQCAAFNS--- 396  
DB 536 CGGVQFSSRDCTFRVPRNGKCYEGRTRFRSCNTEDCPTGSALTFRBEOCAAYNHRTD 595  
QY 397 --OEFMGOLYOWEP--FTEVQSGORCELNCRPRGRFFVYRHTKVDQDGTLCQPGAPDICVA 453  
DB 596 LEKSPFGPK-DWPRYTGAVPDQCKLTCQARALGYVYVLEPRVVDGTGCPSPDSSSVQVQ 654  
QY 454 GCLSPGCDGLGSGRRPDGCGVGGDDSTCLVSGNLTDRGGLGTQKILWIPAGALRL 513  
DB 655 GRCIHAGCDRIIGSKKFKDKMVCSCGSGSGSKGSGP--RKFYRYGNVNVVITPAGATHI 712  
QY 514 QIAOL-RS--SNYLALRPGGRSITINGNAV--DPPGSRAGGTVFYRNPREECKGE 568  
DB 713 LVROQNGPHRSIYLALPLDGSYALNEEYILMPSPTDVLPGAVSLYS---GATRASE 769  
QY 569 SLASRGPTTPQVDVYMI--QEENGVFYQYVYSSPPPILENPTPEPPVPOLQ--EILR 624  
DB 770 TSLGHPGLAQLTLQVLVAGNPQDRLYSFVPRPTSTPRPTQDMLHRRRAQLEILR 829  
QY 625 VEP 627  
DB 830 RRP 832  
RESULT 15  
AT58\_MOUSE  
ID AT58\_MOUSE STANDARD; PRT; 905 AA.  
AC P57110;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-8 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase  
with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).  
GN ADAMTS8  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20079168; PubMed=10610729;  
RA Georgiadis K.E., Hirshata S., Seldin M.F., Apte S.S.;  
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on  
mouse chromosome 9 and human chromosome 11.";  
RL Genomics 62:312-315(1999).  
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART  
AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY  
SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AF175282; AAF25805.1; --  
DR MGD; MGI:1353468; Adamts8.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR001590; Reprolysin.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR000130; Zn.MTpeptdse.  
DR Pfam; PF01421; Reprolysin; 1.  
DR Pfam; PF00090; tspl-1; 2.  
DR SMART; SMO0209; tspl-1; 2.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
DR PROSITE; PS0215; ADAM\_MEPRO; 1.  
DR PROSITE; PS0092; TSPI; 1.  
DR PROSITE; PS00427; DISINTEGRIN-1; FALSE\_NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 28  
FT PROPEP 29 228  
FT CHAIN 229 905  
FT METAL 378 378  
FT ACT\_SITE 379 379  
FT METAL 382 382  
FT METAL 388 388  
FT DOMAIN 453 541  
FT DOMAIN 542 598  
FT DOMAIN 599 705  
FT DOMAIN 706 847  
FT DOMAIN 848 905  
FT CARBOHYD 415 415  
FT CARBOHYD 480 480  
FT CARBOHYD 506 506  
FT CARBOHYD 615 615  
SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;  
Query Match 8.2%; Score 399.5; DB 1; Length 905;  
Best Local Similarity 27.2%; Pred. No. 1.5e-11;  
Matches 123; Conservative 59; Mismatches 161; Indels 109; Gaps 18;  
QY 321 GTPHGR-----LEPDQHP-----GAWLPLLSNGPHASSLSWLPAPS----- 358  
DB 514 GTPCGPHLCLDGSCVLKEDVENKAVVDGWDGWPWP-----WQCSRTCGGI 562  
QY 359 -----SPIPR-----CSGESQLRACSQAPCPPEQPPRALQCA---AFNSQEFMGQ 402  
DB 563 QFSNRECDNPNPQNGRFLCGERVKYQSCNTECPNGKSFREOQCEKYNAYNHTDLGN 622  
QY 403 LYQWEP-FTEVQSGORCELNCRPRGRFFVYRHTKVDQDGTLCQPGAPDICVAGRCLSFGC 461  
DB 623 FLQWPKYSGVSPDRCKLFCRARGRSEFKVFEAKVIDGTILCGPDTLSICVRCQCVKAGC 682  
QY 462 DGIILGSRPPDGCVCVGGDDSTCLVSGNLTDRGGLGTQKILWIPAGALRLQIAQ---- 517  
DB 683 DHVNSPKLKDCKGCVGKGGKACRKISGFTFFS--GYNDIVIPAGATNIDVKQRSH 740  
QY 518 -LRPSSNYLALRPGGRSIIINGNAVDP-PGSRAGGTVFYRNPREECKGESLSAEGP 575  
DB 741 GVRNDGSYALKNTANGQYLLNGNLAIQSL--PSAEWVLGDWSECPSTC-RGSWRRTVEC 797  
QY 576 TTQPDVYMIQEENPGVYQYVYSSPPPILENPTPEPPVPOLQPEILRVEPPPLAPAPRP 635  
DB 798 -----LPEP-----LTVQLTVSGEVP-PKV 818  
QY 636 ART-----PGLQVRQVRPQMPAPPHPRTPLGSPAAYKRWKVSACSACSGKGVWRPIFCI 692  
DB 819 RYTFVPNDMDFSVQNSKERATTNLIQSL--PSAEWVLGDWSECPSTC-RGSWRRTVEC 875  
QY 693 SRESSEELDERSCAAGAPPPASPECHGTGCP 724  
DB 876 RDPGQASD--TCDE-ALKPEDAKPCGSPQCP 904  
Search completed: July 23, 2002, 21:13:24

us-10-041-770-2.rsp

Wed Jul 24 11:39:35 2002

Job time: 478 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 20:20:06 ; Search time 43.09 Seconds  
(without alignments)  
1955.682 Million cell updates/sec

Title: US-10-041-770-2

Perfect score: 4895

Sequence: 1 MENWGRPWLYLLLSLPQ.....PPAISCILGNHAQTSAPPA 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	640.5	13.1	1558	C89114	protein C37C3.6a [
2	640.5	13.1	2167	T34395	hypothetical prote
3	618	12.6	951	T00260	hypothetical prote
4	618	12.6	2165	T21371	hypothetical prote
5	531	10.8	1205	T18517	procollagen N-endo
6	486.5	9.9	951	T00017	gene ADAMTS-1 prot
7	466	9.5	550	T47158	hypothetical prote
8	414.5	8.5	898	T14764	hypothetical prote
9	403.5	8.2	1059	T22545	hypothetical prote
10	399.5	8.2	937	T00355	hypothetical prote
11	284	5.8	788	T25061	hypothetical prote
12	277.5	5.7	2142	B35098	MHC class III hist
13	276.5	5.6	1870	S37671	MHC class III hist
14	276.5	5.6	1872	S36152	MHC class III hist
15	262	5.4	660	Q0B63	BHLF1 protein - hu
16	260	5.3	2715	T213049	eyelid - fruit fly
17	235	5.2	2849	A38587	hypothetical prote
18	235	5.1	1146	A38587	collagen, cornea's
19	250.5	5.1	924	A59266	gene Irf3 protein -
20	248	5.1	3530	A59266	unconventional myo
21	244	5.0	3511	A59295	atrophin-1 - human
22	242.5	5.0	1184	S50832	hypothetical prote
23	238.5	4.9	963	T19140	hypothetical prote
24	237	4.8	957	T15976	hydroxyproline-ric
25	235.5	4.8	620	S06733	brain-specific ang
26	235.5	4.8	1572	T00027	collagen alpha 1(X
27	226.5	4.6	1532	A61262	atrophin-1 - human
28	225.5	4.6	1184	G01763	atrophin-1 - human
29	225.5	4.6	1584	T00026	brain-specific ang

30	224	4.6	1151	2	T18535	high molecular mas
31	222.5	4.5	1444	2	T18856	angiotensin inhib
32	220	4.5	4957	2	T03455	ALR protein - huma
33	220	4.5	5262	2	T03454	ALR protein - huma
34	214.5	4.4	712	2	A45638	immunodominant mic
35	214	4.4	1736	2	T00391	hypothetical prote
36	213.5	4.4	1464	1	CGHU15	collagen alpha 1(I
37	212.5	4.3	1496	1	CGHU2V	collagen alpha 2(V
38	212	4.3	1433	2	A46053	bullous pemphigoid
39	212	4.3	1690	1	CGHU1B	collagen alpha 4(I
40	211	4.3	1546	1	CGHU2E	collagen alpha 2(X
41	210.5	4.3	1049	1	CGBO7S	collagen alpha 1(I
42	210	4.3	1838	1	CGHU1V	collagen alpha 1(X
43	209.5	4.3	1774	2	B56101	collagen alpha 1(X
44	209	4.3	886	2	I50694	collagen alpha 1(I
45	208.5	4.3	753	2	JQ0532	OP protein - Kenne

ALIGNMENTS

RESULT 1

C89114  
protein C37C3.6a [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: C89114  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio  
A:Reference number: A75000; PMID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;  
A:Accession: C89114  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1558 <STO>  
A:Cross-references: GB:Chr\_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CESP:C37C  
C:Genetics:  
A:Gene: C37C3.6a  
A:Map position: 5

Query Match	13.1%	Score	640.5	DB	2	Length	1558
Best Local Similarity	28.0%	Pred. No.	6.7e-24				
Matches	149	Conservative	77	Mismatches	215	Indels	91
Gaps	12						
Qy	344	NGPHASSIWSLFAFSSPIPR-----	CGSEQLRACSOAPCPPEQDPDR	387			
Db	71	SGQKETGNWGPVVENECSSCGGVQLEKQCSGDCTGASVYISCNLNAC-ESGTDFR	129				
Qy	388	ALQCAAFNSQEFMGQLYQWEPFTEVQGSQRCNLNCRGRFFYVYRHTKVKVODGTLQCPGA	447				
Db	130	AEQCKFNDEALDGNVHKWTPY---KGKNKCELVCKPESGNFYFKWADKVVYDGTKCDSKS	186				
Qy	448	PDICVAGRCISPGDGIILSGRRPDGCGVCGDSTCLVSGNLTRDGGPLGYKILWIP	507				
Db	187	NDICVDGCELPVGDCKGLSKFKDCKGCKDGGSTCKTIEGRDERNLSFGYHDIKLP	246				
Qy	508	AGALRLQIAQLRPSNSYLALRGPGRSRINWAVDPGSGYRAGGTVFYRNPREEKG	567				
Db	247	EGATWIKIQEARKSTNNLALANGSDHFLNGLIQVEKEVEVGTFIVY-----DDAEP	301				
Qy	568	BLSAEGTTPQVDVYMFQENPGVFQYVYSSPPPILENTPPEPPVQLOPEILRVEP	627				
Db	302	ETLSAQGLSELTALLFRKGRDRTAIKYEFISIP-----LEEEV-----	341				
Qy	628	PLAPAPRPTPGTLQROVRIPQMPAPPPTPLGSPAAYKRVKRGHSACSACSGKGVWRP	687				
Db	342	-----DYMFKEDNWTPTCSVSCGKGQVTR	364				
Qy	688	IFLCISRSGEELDRSCAA-GARPPASPECHGTPTCPPYWEAGETWTSRSCG-PGQHQH	745				

Db 425 RVYCHQVANGRRVTVEDGNCITVERPPVKTCTNCFACPEWQAG-PWSACSEKCGDAFOY 483  
QY 806 ROVRC-----VGNNGDEVSEQECASGPPQ-PSREACDMGPCITAWF-HSDWS 851  
Db 484 RSVTCRSEKEGEGKLLAADACPADEQERFDTERTCNLGPCEGLTFVTGEWN 535  
RESULT 3  
T00260  
hypothetical protein KIAA0605 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
R:Accession: T00260  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
A:Reference number: Z14086; MUID:98290545  
A:Accession: T00260  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-951 <NAG>  
A:Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BAA25531.1; PID:g3043734  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0605  
C:Superfamily: thrombospondin type 1 repeat homology  
F:46-106/Domain: thrombospondin type 1 repeat homology <THRI>  
Query Match 12.6%; Score 618; DB 2; Length 951;  
Best Local Similarity 25.2%; Pred. No. 5.3e-23;  
Matches 192; Conservative 83; Mismatches 235; Indels 252; Gaps 28;  
QY 319 TGGTTPHGFRLEPPDPHPGAWLPLLSNGPHASSL-----WSLFA----- 356  
Db 27 TGTSTDNSTNS-----LEGGTDATAFWWGWTWKTAFSRSCGGGVTSQERHCL 75  
QY 357 ----PSSPIP---RCSESEQLRACSOAPPEQDPDRAQCQAFNSQEFMGQLYOWEP- 408  
Db 76 QORRKSVPGNRRTCTGTSKRYQLCRVQECPPDGRSFRQCVSFNSHVYNGRTHQKPL 135  
QY 409 ----FTEVQSGQRCELNCRP-RGFRFVVRHTEKVDGTLCO-POAPDICVAGRLSPGCD 462  
Db 136 YPDYVHI-SSKPCDLHCTTVDGQR---QLMVPARDTSCKLTDLRGVGVSGKEPIGCD 191  
QY 453 GILSGRRPDGCGVGGDDSTCLRVSGNLTRDGGPLGYQKILWIPAGALRLOIAQLRPS 522  
Db 192 GVLFTHTLDRKCGICQGGSSCTHTGNYRKGNAHLGYSLVTHIPAGARDIQIVERKKA 251  
QY 523 NYLALRGPGRSIINGWAVDPGSGYRAGTVFVRYNRPREEGKG-ESLSAEGTTPQVD 581  
Db 252 DVLALADAGYFFNGYKVDSPKNFNIAIGVYVYRPRMDVYETGIEVIVAOGPTNOGLN 311  
QY 582 VYMFQE---ENPGVYQVVISPPPILENPTPEP-----VP-- 616  
Db 312 V-MVWQNGKSPSITFEYTLQPP---HESRPQIYYGFSESQSLDGLAGLMGFIPIHN 367  
QY 617 ----OLOPEILREPPILAPAP-----RPARPTGTIQR 644  
Db 368 GSLYQASSERLGLDNLRFHGPGLDMELPGQGOETNEVCQAGGACGPPRGKGRDR 427  
QY 645 QY-----RTPQMPAPHP 657  
Db 428 NYTGPTLTKDDEEDVTHFASQEFSSANAISDQLLGAGSLDKDFTLNETVNSIFAQAP 487  
QY 658 RTPIG----- 662  
Db 488 RSLAESFVDEEYNEGAGPYLLNGSYLELSSDRVANSSSEAPPNVSTSLISAGNRTH 547  
QY 663 -----SPAAY--WKRVGHSAACSCGKGVWRPFLCISRESGELDERSCAA 707  
Db 548 KARTRPKARKQGVSPADMYRWKLSLSHEPCATCTTGVMSAYAMCV-RYDGVVEDDSYCYDA 606

Db 365 NLYCIDGNKGRVEDDICEENNAATKPEFEKSCETVDCAEAWFTGDWBSCSSTCGDQGOQY 424  
QY 746 ROLOCRQEFGGGSSVPPERCGLHPRNITQSCQLRCLGHEWSPSQSVRCGRGRQS 805  
Db 425 RVYCHQVANGRRVTVEDGNCITVERPPVKTCTNCFACPEWQAG-PWSACSEKCGDAFOY 483  
QY 806 ROVRC-----VGNNGDEVSEQECASGPPQ-PSREACDMGPCITAWF-HSDWS 851  
Db 484 RSVTCRSEKEGEGKLLAADACPADEQERFDTERTCNLGPCEGLTFVTGEWN 535  
RESULT 2  
T34395  
hypothetical protein C37C3.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
A:Accession: T34395; T34394  
R:Geisel, C.; Bradshaw, H.  
submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of C. elegans cosmid C37C3.  
A:Reference number: Z21518  
A:Accession: T34395  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2167 <GEI>  
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GNO0023; CESP:C37C3.6b  
A:Experimental source: strain Bristol N2; clone C37C3  
A:Accession: T34394  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1555, 'SKF' <GE2>  
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GNO0023; CESP:C37C3.6a  
A:Experimental source: strain Bristol N2; clone C37C3  
C:Genetics:  
A:Gene: CESP:C37C3.6b; CESP:C37C3.6a  
A:Map position: 5  
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 1556/  
Query Match 13.1%; Score 640.5; DB 2; Length 2167;  
Best Local Similarity 28.0%; Pred. No. 9.1e-24;  
Matches 149; Conservative 77; Mismatches 215; Indels 91; Gaps 12;  
QY 344 NGPHASLSLFASSPIPR-----CSGESQLRACSOAPPEQDPDRAQCQAFNSQEFMGQLYOWEP- 387  
Db 71 SGQKGTNGPWPVNECSRSGGGVQLEKQCSGDCTGASVRYTSCNLNAC-ESGTDPR 129  
QY 388 ALOCAAFNSQEFMGQLYOWEPFTEVQSGORCELNCRPGRFVVRHTEKVDGTLCOGCA 447  
Db 130 AEQCSKFNDALDGNHYKWTY---KGKNCCLVCPESGNFYKWKADKVDGTKCDKSKS 186  
QY 448 PDCVAGRLSPGCDGILSGRRPDGCGVGGDDSTCLRVSGNLTRDGGPLGYQKILWIP 507  
Db 187 NDICVDECLPVGCDKGLSSLKFKCKGCDGSGTCKTIEGRFDERNLSPGYHDIKLP 246  
QY 508 AGALRLOIAQLRPSNYLALRGPGRSIINGWAVDPGSGYRAGTVFVRYNRPREEGKG 567  
Db 247 EGATNRIQIARKSTNLNALKNGSDHFLYNGSLIGLQVEKEVEGGTIFYI-----DDAEP 301  
QY 568 ELSAEGTTPQVDVYMFQENPGVYQVVISPPPILENPTPEPVPQLOPEILRVERP 627  
Db 302 ETLAQGPLSEETVALLFKGSRDTAIKYEFSIP-----LBEV----- 341  
QY 628 PLAPAPARTPTGLQQRVIRPOMAPPHPRTPLGSPAAVYKVKVGHSAACSCGKGVWRP 687  
Db 342 -----DYMKFEDNWTPCSVSGKGVQTR 364  
QY 688 IFLCISRESGELDERSCAA-GARPPASPEFCHTGPVYWEAGEWTSRSRSG-PGTQH 745  
Db 365 NLYCIDGNKGRVEDDICEENNAATKPEFEKSCETVDCAEAWFTGDWBSCSSTCGDQGOQY 424  
QY 746 RLOCRQEFGGGSSVPPERCGLHPRNITQSCQLRCLGHEWSPSQSVRCGRGRQS 805

```

QY      316 PWGTGGTP-----HGP--RLEPD--POHPGAMLPLLNSNGPHASSLM-----SLFA   356
D5      571 FW-ADGTPCDESRSNFCHHGACVRLAPESLTKIDOGMDWRMSWGECSNTCGGVQKGLRD   629
QY      357 PSSPIPR-----CSGESBQLRACSQAOPCPPEQPDPRALQCAAFNSQBF-----MGQLYQW   406
D5      630 CDSKPXRGNGKYCVGOREYRSCHTQECPWDTQYREVQCSEFNKKDIGIQGVASTNTHW   689
QY      407 EP-FTEVGOSORCELNCRPGRFRFYVRHTEKVQDGTTCCQPGAPDICVAGRCLSFGCGIL   465
D5      690 VPKYANVAPNERCKLYCRLSGSAAFYLLDRXVDVGTPCDRNGDDICVAGACMPAGCDHQL   749
QY      466 GSRERPDGCGVCGGDSTCLRVSGNLTRDGGPLGYQKLIWPAGALRLQIAQ-----LRP   520
D5      750 HSTLRDKCGVCGGDDSSCKVKVTNEQ-GTFGYNEVMKIPAGSANIDIRQCYNNMKE   808
QY      521 SSNYIALRGPPGRSRIINGNAVDDPGSYRA-GGVFVRNRPPRECGKGESLSASGPTQP   579
D5      809 DDNYLSLAANGNEFLNHGFVSLARQIAFDQTVEYS---GSDAIFERINGTGPIRSD   865
QY      580 VDVMYI-FQEENPGVFYGVYTSPPPILENTPEPPVPOLQPEILRVEPPLAPAPRAPRT   638
D5      866 IYHVLSVGSHPDLSYEWMTAAVD---NAVIRPISALY--LWRVTDTWTECDRACRG   919
QY      639 P-----GIQR   644

```

Query Match	10.8%	Score 531;	DB 2;	Length 1205;
Best Local Similarity	26.2%;	Pred. No. 1.1e-18;		
Matches 164; Conservative	70;	Mismatches 235;	Indels 156;	Gaps
Qy	289	WASPVQAGRDPDPFSPVPRGRGOQGGPMGTGTGTHGPRLPEDPORHGAWLPLLS-----	343	
Db	546	WLTPDLIKR-----DGNM-----	571	
Qy	344	-----NGPHASSLWSLEAFSSPIPRGSGESELACSAQCPPEQDPDRALOCA	392	
Db	572	GTGWKFTTRCDMPH-----PANGRTCSGLAYDQLCNSDCPDALADFFEEQCR	622	
Qy	393	AFNSQ-EFMGQLYQWEPFTEVQSGSCELNCRPGRFRFVYRHTKVDQTLQ-Q-EGAPDI	450	
		:::	:::	
Db	623	QWDLYFEHGDQAQHWLPHEDHDAKERCHLYCESKETGEVYVSMKRVHDGTRCSYKDAFSL	682	
		:::	:::	
Qy	451	CVAGRCLSPGCDGTILSGRRPDCGCGVGGDDSPICRLVSGNLDRGGPLGYQKILWPAGA	510	
		:::	:::	
Db	683	CVRGDKRVKVGCDGVIGSGSQEDKCGVGGGDSHCKVYKFTFSRPAKGLGYIKMFIPAGA	742	
		:::	:::	
Qy	511	LRLQIAQLRPSNYALURG-PGGRSIIINGMAYDPFG-SYRAGGTVFYRNRPPEKGGKE	568	
		:::	:::	
Db	743	RHLIIQBADTTSRHLAVKNLETCKFIILNEENDVDPNKSFTIAMGVWEWY-----RD	798	
		:::	:::	
Qy	569	SLSREGTTPQVDVYIMTFQ-EENPGVGYQYVSSPP-----PILENPTPEVPVQLQPE	621	
		:::	:::	

```

Db 799 TLQTMPLHGTITVLVPEGDARISLTKYMIHEDSLNVDNNVLED----- 845
QY 622 ILRVEPLAPAPRARTGTGLQVRIPQMPAPPHRTPLGSPAAYKRVKRGHSACSASCG 681
Db 846 -----DSVGEWALKWKSCKPCG 865
QY 682 KGVNRPFLICISRESGEELDERSCAAGARPASPCHGTPC-PPYWEAGBWTSCRSRSG 740
Db 866 GGSQFTKYGCRRLDHKVMHVRGDCSVSKPAIRTCNPQCECQSPVWVTGEWEPSCRSRG 925
QY 741 -PGTHROLOCRQEGGGGS-SVPERCGHLPARNITQSCOLRLC-GHWEVGSFWSQSV 797
Db 926 RTGMQVRSVRCVQPLHNNNTTSVTHKND-ARPEGRACNRELCPGRWRAGS-WSQCSV 983
QY 798 RCGRQSRQVRC-VGNNGDEVSEOCASGPPQPSREACDMGPC-----TTAWFHS 849
Db 984 TCGNTQERPVLCRTADDSFGVCREE-----RPEARICRLGPCRNTSDPSKSYVQ 1037
QY 850 WSKVSPPEPPALCILGHNADTS 874
Db 1038 WLSRPDPNSP-----VQETSS 1053

RESULT 6
T00017
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017
R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.
Genomics 46, 466-471, 1997
A>Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene
A:Reference number: Z14035; MUID:98110583
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-951 <KUN>
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057
A:Experimental source: strain 129SVJ
C:Genetics:
A:Gene: ADAMTS-1
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2
C:Superfamily: thrombospondin type 1 repeat homology
F:542-598/Domain: thrombospondin type 1 repeat homology <THR>

Query Match 9.9%; Score 486.5; DB 2; Length 951;
Best Local Similarity 23.7%; Pred. No. 1.3e-16;
Matches 149; Conservative 74; Mismatches 186; Indels 219; Gaps 24;

QY 250 RHHPRAQAGTEPPSTHSLGSGGFFRASQPQPRPSSQGWASQVAGRRPDPFPSPVPRGR 309
Db 508 KHPF-WADGT-----SCGEGKVCVSKCNKTKDKHFATPV-----H 543
QY 310 GOQGO-GPMGT-----GGTPHGRLEPDQHPHGAWLPLLSNGPHASSLSLAFAPSSIP 362
Db 544 GSWGPMGPGDCSRTCGGVQVYTM-RCD-----NEVP 575
QY 363 R-----CSGESQLRACSOAPCPBPQDP-RALQCAAFN--SQEPMQO--LYOWEP-FTE 411
Db 576 KNGKYCBGRKRVYRSCNTECDPNNKTFREECEAHNEFSKASFGNEPVETPKYAG 635
QY 412 VQGSORCELCNRPGRFRFYVHTEKRVQDGLCOQGPAPICVAGRCLSPGCDGILSGRRP 471
Db 636 VSPKDRCKLTCAKGIYGFVLPQKVVYDGTGCPSPDSTSVCGQCVKAGCDRIIDSKKFF 695
QY 472 DCGCVGGDDSTCLVSGNLTDRGGPLGYQKILWIPAGALRLQIAQ-----LRPSSNYLA 526
Db 696 DKCGVGGNGSKCKKSGIVTSTRP--GYHDIIVTIPAGATNIEVHRNQRNNGSFLA 753
QY 527 LRGPGRSRIINGNW---AYDPPGSRAGTGVFRYRNPREEKGESLSAE-----GPTTQ 578
Db 754 IRAADGTIYLNFTLTLEQDLTYK--GTVLRS-----GSSAALERIRSFSPKLE 803
```

```

QY 579 P--VDVMIFOEBNPGVYQYVVISPPILENTPEPPVQLOPEILRVEPPLAPAPRA 636
Db 804 PLTIQVLWGHALRPKIKTYFMKK----- 828
QY 637 RTPGTLQVRIPQMPAPPHRTPLGSPAAYKRVKRGHSACSASCGKGVWRPIFLCISRES 696
Db 829 ----- 828
QY 697 GEELDERSCAAGARPASPCHGTPCPCPPYWEAGBWTSCRSRSGTQHRLOQCRQEFEG 756
Db 829 -----KTESFNAITPFSWIEENGESKTCGSGWRVQCRDINGH 871
QY 757 GGSVPPPERCGHLPARNITQSCOLRLCGRHWEVGSNPSQSVRCGRQSRQVRCVGNNGD 816
Db 872 -----PASECAKEVKPASTRPCADLPCPHQVGD-WSPGSKTCGKGYKKRTLKCVSHDGG 925
QY 817 EVSEQECASGPPQPSR--EACDMGPCT 842
Db 926 VLSNESC--DPLKKPKHYIDFCLTQCS 951

RESULT 7
T47158
hypothetical protein DKFZp762C1110.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
A:Experimental source: adult melanoma (Mewo cell line); clone DKFz762C1110
C:Genetics:
A:Note: DKFZp762C1110.1
```

Query Match 9.5%; Score 466; DB 2; Length 550;  
Best Local Similarity 22.9%; Pred. No. 7.5e-16;  
Matches 136; Conservative 71; Mismatches 179; Indels 208; Gaps 22;

```

QY 318 GTGG-----TPHGRLEPDQHPHGAWLPLLSNGP-----HASSLSLFLAP 357
Db 96 GTSGLVLCOTKHHPWADGTSCEGKW---CINGKCNKTKDRKHFDTPFHGS--WGMWG 150
QY 358 -----SSPIPR-----CSGESEQLRACSOAPCPBPQDP-RALQCA 392
Db 151 WDCSRTCGGVQVYTMRECDNVPVKNKGKCEGKRVYRSCNLEDCPDNNGKTFREEQCE 210
QY 393 AFNSQEF-----MGQLYQWEP-FTEVQGSORCELCNRPGRFRFYVHTEKVDGTLQCP 445
Db 211 AHN--EFSKASFGGPAVEWIPKYGVSFDRKDLCOAKGIGYFFVLPQKVVYDGTFCSP 268
QY 446 GAPDICVAGRCLSPGCDGILSGRRPDPGCGVGGDDSTCLVSGNLTDRGGPLGYQKILW 505
Db 269 DSTSVCGQCVKAGCDRIIDSKKFKDKCGVGGNGSTCKKISGVT--SAKPGYHDIIT 326
QY 506 IPAGALRLQIAQ-----LRPSSNYLALRGPGRSIINGNW---AVDPPGSRAGTGVFR 557
Db 327 IPTGATNIEVHRNQRNNGSFLAKADGTVILNGDYTLTLEQDIIWK--GVVLY 384
QY 558 NRPREEGKESLSAE-----GPTTQP--VDVYMIFOEBNPGVYQYVVISPPILENPT 610
Db 385 S-----GSSAALERIRSFSPKLEPLTIQVLTGVALRPKIKTYFMKK----- 427
QY 611 PEPPVQLOPEILRVEPPLAPAPPHRTPLGSPAAYKRVKRGHSACSASCGKGVWRPIFLCISRES 670
Db 428 ----- 427
```

```
QY 671 VGHSAASACGKGVWRPFLICISRESGEELDERSCAAGARPASPPECHGTPCPPIWEAG 730
Db 428 -----KKESENAITFTSAMWIE 444
QY 731 EWTSCSRGCGTQHOLQCRQEFGGSSVPPERCCHLPRNITQSCOLRLCGHWEVGS 790
Db 445 EWGCKSCGELGWRRIVECRDING-----QPASECAKEVPASTRCADHPCPQWOLGE 499
QY 791 PWSQSVRCGRQRQRVCGVNGNGDEVSECECAGPPQPPSR--EACDMGPGCT 842
Db 500 -WSSCKTCGKGYKRSKLCLSHDGVLSHESC--DPLKPKHFDICTMAECS 550

RESULT 8
Tl4764
Hypothetical protein DKFp434H204.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: Tl4764
R:Wanburt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18181
A:Accession: Tl4764
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-898 <WAM>
A:Cross-references: EMBL:AL110226
A:Experimental source: adult testis; clone DKFp434H204
A:Genetics:
A:Note: DKFp434H204.1

Query Match 8.5%; Score 414.5; DB 2; Length 898;
Best Local Similarity 24.8%; Pred. No. 3.7e-13;
Matches 234; Conservative 63; Mismatches 311; Indels 337; Gaps 53;

QY 19 PQLCLDQVLSGHSLOTPTRECGQEGVGVWQVWASQSCQCGVQVRSRTC----- 71
Db 133 PRAC-----EHLPPPTETPCNRHVPCPATWAVG-NWSQCSVTGEGTQRNVLCTINDTQVP 188
QY 72 ---QLPTVOLHPSLPLPPRPHPEALL-PRGQGRPRTSPETPLRYTQSRGGGLR 126
Db 189 CDEAQAQPASEVTCISPL---CRWPLCLTLPESG-----S 220
QY 127 GPASHLGREETOET---RAARSRRLDPTKPMFGYGRVFPFALPLHRNRHPRSPREL 183
Db 221 GSSHELFEADFIHPLAPRSPASSP-KPGTMG-----NAIEEAP---EL 264
QY 184 SLISSRGEAIPST-----PRAEPFSANGSPOTELPPTLSVHTPSQAERPL 231
Db 265 DL-----PGPVVDFFDYDYNFINHEDLSTGSEEDLDLAGTGDRTTPPHSRPA 315
QY 232 SPETAQTEVAPRTAPALRHHRAQASGTPTPSHSLGEGFFRASFPQPRPSQGWAS 291
Db 316 APSTGS-----PVP-----ATEPPAAKE---EGVLGPWSPSP-----W-- 345
QY 292 PQVAGRRPDPFSPVPRGCGQGGWGTGGTGHGPRLEDPDH-----PGAWLPILLSNG 345
Db 346 PSQAGRSPPP-PS-----EQ-----TPGNFLINFLPEEDTPIGADPLGLSL-S-W 388
QY 346 PHASSLSLFAFSPSPTRCSESEQ---LRACSAQPCPEQDPRALQCAAFNSQEFMGQ 402
Db 389 PRYST-DGLQTPATP-----ESQNDFFVKGDSQSLPPPPWRDRTN---EVFKDDE--- 434
QY 403 LYONWPFTEVQGSQRCLENCRRGRFRFYRHTEKVDQGLTQOPGADICVAGRCISPGCD 462
Db 435 -----EPKGRGAPHLPRP-----SSTL-----PP-----LSP--V 458
QY 463 GILGSRPDCGV-CGGDDSTCLRYSGNLTDRGGPLGYOKILW-----IPA 508
Db 459 GSTHSFSPDVAELWTGGTVAVPEALEGL-----GPVDS--LWPTVGVASLLPPIAPL 512
QY 509 GALLRLQIAQLRPSNNYLALRGPGRSIIN-GNMAVDPPGSYRAGGTVFYRNRPPREGKG 567
```

```
Db 513 PEMKVRDSSLEPGTSPFPAPGCSWDLQTVAVWGTFLPTILTGLGHMPEPALNPCKGQP 572
QY 568 ELSAEGTTTPQVDVYVMIFOENPGVYQVVISPPILENPTPEPPVQLOQEILRVEP 627
Db 573 ELSPEVPLS-----SRLST--PAWDSANSHRVPETQP---LAP 608
QY 628 PLAPAPAPARTPGTLQQRVIRPQMPAPPHPTPLGSPAAVYKRVGHSACSASCGKG-VWR 686
Db 609 SLAEGPPA-----DPLVYNASWQAGNWSSECTTTCGLGAVMR 646
QY 687 PIFLCISRESGEELDERSCAAGARP-PASPEPCHGTPCPPIWEAGWTSRSCGPGTQH 745
Db 647 PV-RC-----SSGRDED---CAPAGRPOPA--RRCHLRPCAT-WHSGNWSKSCRSCTGGSSV 696
QY 746 RLQLOC-----RQEFGGG----- 758
Db 697 RDQVCYDTRDLRLRPLRPHCPGPAKPAHPCGAQPCLSWYTSWRECSACCGGQQRL 756
QY 759 -SSVPPERCCHLPRNITQSCOLRLCGHWEVGSVWSCSVRCGRGORSRQVRCV----- 811
Db 757 VTCPEPGLCEALRPNTTRPCNTHPCQTQWVVG-PWGCSAPCGGVQRRVLKCVNTQIGL 815
QY 812 -----GNNGDEVSECECAS---GPPQPP--SREACDMGPGCTT 843
Db 816 PEEDSDQCGHEANPESSRPGCTDECEVPEPPRCERDLRSFGCET 860

RESULT 9
T22545
Hypothetical protein F53B6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22545
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19578
A:Accession: T22545
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1059 <WIL>
A:Cross-references: EMBL:Z81086; PIDN:CA803121.1; GSPDB:GN00019; CESP:F53B6.2
A:Experimental source: clone F53B6
A:Genetics:
A:Gene: CESP:F53B6.2
A:Map position: 1
A:Introns: 38/3; 92/3; 131/3; 169/1; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3;

Query Match 8.2%; Score 403.5; DB 2; Length 1059;
Best Local Similarity 22.9%; Pred. No. 1.5e-12;
Matches 121; Conservative 42; Mismatches 151; Indels 215; Gaps 16;

QY 347 HASSL-WSLFAPSP-----IPRCGESQLRACSAQPCPPQPDRA 388
Db 20 HVDALSWAASFPWSSCTKTCGGVSRQLRCLTSKCSGESYRFKVCQAKTCESRLRAD 79
QY 389 LQCAAFNSQEFMGQLYQWEPTEVQGSQRCLENCRRP--GFRFYVRHTEKVDQGLTQCPG 446
Db 80 TICGG-----EEIVSRQCEVCSRILTGANFLWR---VDDGTQCOAA 119
QY 447 -APDTCVAGRCLSPECDGILSGRRPDCGCGVGGDDSTCLRYSGNLTDRGGPLGYOKILW 505
Db 120 TSRAVCSKSGCOIVGCDGLISSFRFDACGVCGRGDTG-----DNG-----RFLW 165
QY 506 IPAGALRLQIAQLRPSNNYLALRGPGRSIINGWAVDPPGSYRAGGTVFYRNRPPREG 565
Db 166 -----KVSEETACAS-----NCDDIVD----- 183
QY 566 KGESLSAEGTTPQVDVYVMIFOENPGVYQVVISPPILENPTPEPPVQLOQEILR 625
Db 184 ----- 183
```

QY 626 EPPLAPAPRARTPCTLRQVRIPQMPAPHPRTPLGSPAAWKKRVGHSHSACSGKGVW 685  
Db 184 -----WSGAGRSIASTS----- 195  
QY 686 RPIFLCISRESGELDERSCAAGARPPASPEFCHGTGTPCPYWEAGETWSCSSCGPGTQH 745  
Db 196 QPIVVCVNAITGRVWPEKLCADLRKVEARPCPMLICPSRWMAADWTECVPHCGEGTRK 255  
QY 746 ROLOCQERGGSSVPPERCCHLPRNITQCOLRLCHGHEVSPWQCSVRCRGGRS 805  
Db 256 REVICVOTAHNVTHVDFTEFCNGTTPAAEENCVSTCGRWENGK-WSKCTASCQGVRR 314  
QY 806 ROYRCVGNNGNDEYSEOCASGPPQPPSREACDMG-PCTTAMFHSWSSK 853  
Db 315 RHVACVGGSS-----DCDEG-GRPQETTCVAGIPCSSTATSLDNDR 355  
RESULT 10  
T00355  
hypothetical protein KIAA0688 - human  
C:Species: Homo sapiens (man)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
R:Accession: T00355  
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.  
DNA Res. 5, 169-176, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
A:Reference number: Z14142; MUID:98403880  
A:Accession: T00355  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-837 <ISH>  
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BA031663.1; PID:g3327190  
A:Experimental source: brain  
C:Genetics:  
A:Gene: KIAA0688  
C:Superfamily: thrombospondin type 1 repeat homology  
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 8.2%; Score 399.5; DB 2; Length 837;  
Best Local Similarity 25.3%; Pred. No. 1,8e-12;  
Matches 168; Conservative 70; Mismatches 264; Indels 151; Gaps 35;  
QY 61 GVGVRSSRTCOLPTVOL--HPSLPLPPRPHPEALLPR-----GQGR-PQTSPP---E 109  
Db 235 GAGLKRYLLTVMAAAKAFKPSI-----RNPVSLVTRVLVILSGEGPQVGPSAAQ 287  
QY 110 TLPLYRTQSRGRGLRGPASHLG-----RETEQIRAAARSRLRDPKPGMFGYGRVP 163  
Db 288 TLRSCAWQGLNTPEDSDPHEDFTAILTRQDLGCVSTC-----DYL--GMADYGV- 338  
QY 164 FALPLHNRHRRSPRSELSISRG--EATPSPTPRAEPFSA---NGSPOTEL-PPTTE 218  
Db 339 -----CDPARSCAIVEDDGLQSAFTAALHGLHVFNMLHNSKPCISLNGPLS 385  
QY 219 LSVHTSPQAEPLSPETAQTEVAPTRPAPLRHHPRAQASG-----TEPPSPTH----- 267  
Db 386 TSHVWAPVMAHYDPEPNS-----PCSAFITDFLDNGYCHLLDKPEAPLHPVTF 438  
QY 268 -----SLGEGGFFRASQPRRPSQGWASQVAGRRPDPFPSPVPRGQGGGP 316  
Db 439 PGKDYDADROCOITFGPDSRHCPQLPPPCALWCSCGLNG-----HAMCQTKSP 488  
QY 317 WGTGGTTPHGR-----LEPD-----POHPGAWLLLSNGPHASSLSLFPSS-- 359  
Db 489 W-ADGFCGPAQACMGRLHMLQDLQFNIPQ-AGW-----GP-----WGPWDCST 535  
QY 360 -----PIPR-----CSGESBOLRACSOAPCPPEQP--DPRALOCAPNS--- 396  
Db 536 CGGCVQFSSRDCTRPVPRNGGKYCEGRRTRFSCNTEDCPTGSALTFFREOCCAAYNHRTD 595  
QY 397 --OEEFGQLQWEP-FTEVQGSORCELCNCRPGFRFYVHTTEKVODGTGILCOGADICVA 453

Db 596 LKSFQPGPM-DWVPRYTGVAPOQCKLTQCARALGYVYVLEPRVDGTGTCSPDSSSVQV 654  
QY 454 GRCLSPGCDGILSGRRPDCGVCVGGDDSTCLVSGNLTDRGGPLGYQKILWIPAGALRL 513  
Db 655 GRCHAGCDRIIGSKKFKDKCMVCGGDSGCKSQSGSF--RKFYGYNNVVTIPAGATHI 712  
QY 514 QIAQL-RPS--SNYLALRPGGRSIIINGWAV--DPPGYSRAGGTVERVNRPREEGKE 568  
Db 713 LVKQGNFGRHSIYLALKLPDGSYALNGEYTLMPSETDVVLPGAVSLRYS---GATAASE 769  
QY 569 SLNAGETTPQVDVYMF--QENPGVFTYVYVSSPPPLENTPEPPVPQLOP--EILR 624  
Db 770 TLSGHGFLAQPLTLQVLVAGNPQDRLRYSFVPRPTPSTPRPTQDMLHRAQILEILR 829  
QY 625 VEP 627  
Db 830 RRP 832  
RESULT 11  
T25061  
hypothetical protein T21B6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
R:Accession: T25061  
R:Cottage, A.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z19975  
A:Accession: T25061  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-788 <WIL>  
A:Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3  
A:Experimental source: clone T21B6  
C:Genetics:  
A:Gene: CESP:T21B6.3  
A:Map position: X  
A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2  
Query Match 5.8%; Score 284; DB 2; Length 788;  
Best Local Similarity 20.2%; Pred. No. 7e-07;  
Matches 188; Conservative 76; Mismatches 296; Indels 370; Gaps 43;  
QY 15 LLSLPOLCLDQEVLSGSLQTPTTEG-----QGPEGV--WGPWVQWASC- 56  
Db 115 IVRMLELVDCVCHQFDAH---PPLEEDGLASSTDVKVRKQCGDEAVGVDWTDWHS 171  
QY 57 -----SOPCGVGVORRSTCOLPTVOLHPSLPLP-----PRPRHPEALLPRGGR 103  
Db 172 SNHEVRSQACEYGRKIQRGC--PARSAPQVPAPAOQYAPRAPEYPSAQOQQOQREQ 229  
QY 104 PQTSPTPLPLYRTQSRGRGLRGPASHLGREETQEIARAARSRLRDPKPGMFGYGRVP 163  
Db 230 QQ-----RQQH-----REHQARLQHQOQQOQQOQQOQQOQQOQPP-----Op 264  
QY 164 FALPLHNRHRRSPRS-----ELSLSSGEEATPSTPRAEPFSAANGSPQTELPTE 218  
Db 265 QPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 324  
QY 219 LSVHTSPQAEPLSPETAQTEVAPTRP-----APLRHHP-----RAQASGTEPPS 264  
Db 325 FGOSQIQLOSGPVPPOQHPOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 384  
QY 265 PTHSLGEGG--FRASQPRRPSQGWASQVAGRRPDPFPSPVPRGQGGGPWGTGG 321  
Db 385 PKADPCGGCAFPVQAPQOQERTPPPPVPLAPINTATQPLFQ----- 427  
QY 322 TPGHRLPDPQHPGAW-----LPLLSNGPHASSLSLFPSSPIPRCS-----GESE 369  
Db 428 -PYPTRYRPPAPPPACDGGCVNPPVWGVHWDSDWS-----TCSTCTGDGAKS 477  
QY 370 QLRACSOAPCPPEQDPRALQCAAFNSQEFMGOLYQWEPFTEVQGSORCELCNCRGRFR 429

```

Db 478 RRRCSNNC-----OGADYETP-----CNLG-----500
QY 430 YVRHTEKVDGTLQCPGADICVAGRCISPGCDGILGSGRRPD-----GCGVCGDDST 483
Db 501 -----PCQWTS--EWCEWSTC--SASC-----GSGQRERTRCHLGTNRCEGKDYE 542
QY 484 CRLVSGNLTDEGGPLGQKILWTPAGALRLQIAQLRPSSNYLALRPGGSRGSIINGNAVD 543
Db 543 SEQCS-----AGPCP--ENQWEDWQCSTCGQ-----GVAVR--QRTCLGG-----581
QY 544 PPGSYRAGGVFRYRNPREEGSGESLSAEGTTPQVDVYMFQENPVGFFYYVLSPP 603
Db 582 -----VF-----GDHL-CQGPKTE-----QRACDGP 602
QY 604 PILENTPPEVPQLOPEILRVEPIAPAPRPAETCTLQVRIRPQMPAPHPRTPLGS 663
Db 603 CSLWSPQOE-----611
QY 664 PAAWKRVGHSAASACGKGVWRPIFLCISRESGEELDRSCAAGA-----RPPASPEPCHG 720
Db 612 ---W-----STCSASCGSMKR-----RQVCOFGIDCQGNESQFCYG 648
QY 721 TPCPYWEAEWSTCSRSRSGFGTQHRLQCRQEFGGGSSVPPERCCHLPRNI--TQSCQ 779
Db 649 PPCAETWECWESCSCSKCGFGQRTTRGC---LGPNG-----QEATCCQPSIETLCE 700
QY 780 LRLAGHVEGSPWQSCSVRCGRGQ-----RSRQVRCVG 812
Db 701 GQSCCNWSEWCHWCMDCGCGGVRYEYMTGCEWSPCSQLACEVGVQSRQCVG 760
QY 813 NNG-----DEVSEQ-----ECASGPPQPP 831
Db 761 ESGCHCGLAESQCCGLTQC---PPXPP 787

```

## RESULT 12

MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - human  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Aug-1990 #sequence\_revision 06-Nov-1992 #text\_change 24-Aug-2001  
 C:Accession: B35098  
 R:Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990  
 A:Title: A gene pair from the human major histocompatibility complex encodes large pro-  
 A:Reference number: A35098; MUID:90192810  
 A:Accession: B35098  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2142 <BAN>  
 A:Cross-references: GB:M33509; NID:g179338; PIDN:AAA35585.1; PID:g179339; GB:M31293  
 A:Note: the authors translated the codon AGT for residue 97 as Gly  
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 5.7%; Score 277.5; DB 2; Length 2142;  
 Best Local Similarity 22.3%; Pred. No. 3.6e-06;

Matches 226; Conservative 78; Mismatches 359; Indels 351; Gaps 53;

```

QY 24 DQEVLSGH-SLOTPTERGGPE--GYWG-----PWQWASCSQPCGVGVQRRSRTC 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 DEGAEGHRDSQASGSEERPEADGKGNSEPEPTPKTAWAETSRP-----398
QY 72 QLPIVQIHPSLPLPPRP--RHPEALLPRGCGPRQSPETPLPYRQSRGGRGPIRGPS 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 --PTEGPPAPKPLPPGPDYDRGGPPCKPAPADEANQRQKQSSEISLAVERAR 456
QY 131 HLGREETQEIARRSLRDPRIKPMFGYGRVFPALFLHNRNRHPRSPRSELSISSRG 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 RREERERMQEERRAACAEKLC-----RL 481
QY 191 EEATPSTPR--ABPPSANGSPQTELPPTLSVHTSPQAEPLSPETAQTEVAPRTREP 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 482 DEKFGADKRLKAEPAAAPASTPAPPPAVPKELFAPPAP--PASAPT---PETEPEE 536
QY 249 LRHHPRQAASGTE--PPSPTHSLCEG-----GFTRASPT-QPRRPSSQGWASPOVAGR 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 PAQAPPAQSTPTPCVAAAPTIVSGGSTSTSSGSEASVEPOLSKGPEPEPEV--P 594
QY 299 PDPPSPYPR-----GRGQOQGP-WGTGTGTPHGRLEPDPQHPGAWLPLLSNGPHASSL 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 PPTTPPVKVPKGDGIGTTRQPPSOGLGYPKYQKSLPPRFQRCQEQEOLLKQOQHQ--- 651
QY 352 WSLF-----APSSPIPRCSGESEQLRACSQAPCPPEQDPDRALQCAAFNSQEPFGOLYQWE 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 WQHQHQSAPPTPVPPSPQPVLTGAVPAQAPP--PPPKAL-----YPGALGRPP 700
QY 408 PFTVEQSGRCELNCRPRGRFR--YVRHTEKVGQGLTQCPGADICVAGRCISLPGCDGIL 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 PMP-----PMNFDPRMMIPYV-----DPRLLQ--GRPPL---EFYPP---GVH 737
QY 466 GSG-----RRPDGCGVCGDDSTCRLVSGNLTDRGGLGYQKILWI-----PA-----508
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 PSGLVPRERSDSLGL--SSEPFDRHAPAMLRERGTTPVDPKLAWGDVFTATPAERPLT 795
QY 509 -----GALRLQIAQLRPSNYLA-----LRGPGGRSIING 538
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 SPLRQAADDDKGRSETPPVPPPPPYLASYPGPENGAPGPPISRPLEEPGRPL---852
QY 539 NWAYDPPGSYRAGGVFRYRNP--REEGKGESLSAEGPTT-----577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 853 PW---PPGS---DEVAKIQTPPPKPEKETAQLTGPEAGRKLPAASGAGPPPPRE 905
QY 578 -----QPVDYVMIFQENPGVYQYVSSPPPLENPTPEPPVP-----QLQPEILR 824
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 SRTETRMGPRGPRSSRRRGIPEEPGA-----PPRRAGPIKKPPPTKVLELPKPLE 956
QY 625 V--EPFLAPAPRPAR--TPGTQLQVRIPQMPAPHPRTPLG--SPAA-----YMKRVGH 673
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 QGDETTPKPPKPDPLKITGKL-----GGPKETPPNGNLSAPRLRRDYSEYRGP 1006
QY 674 SACSASCGKGVWRPIFLCISR-----ESGEELDRSCAAGA---710
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1007 TSC-----RGRGEXFARGRFGYVGRGRGQGANSAVTESFEEMMGVVEVGGDQTL 1061
QY 711 --PPASPEPCRGTPCPPIWEAGETWTSRSCGPGTQHRQLQCRQEFGGGSSVPPRCGH 768
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1062 LLEAAMPARHGA-----RVQSMKSPSGAGSGAKQAA--1095
QY 769 LPRP-NITQSCQLRLOGHVEGSPWQSCSVRCGRGQSR-----QVRCVGNNGDEV 818
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1096 --RPMRVINLQTRRLPH-----PRRHSRPSRSPTTTRSLHRAPARFTCPGVGESL 1148
QY 819 SEQECASG-----PPO-----PPSREACDMGCTTAWFHSWSSKVSPEPPA 860
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1149 PEGAISGPPRRRREAPQVCPGWSPPAKSLAPKKPPT-----GPLPPS 1190

```

## RESULT 13

S37671  
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 15-Sep-2000  
 C:Accession: S37671  
 R:Bougueleret, L.  
 submitted to the EMBL Data Library, August 1992  
 A:Reference number: S37671  
 A:Accession: S37671  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1870 <BOU>  
 A:Cross-references: EMBL:Z15025; NID:g29374; PID:g29375  
 C:Genetics:  
 A:Map position: 6p21.3  
 A:Introns: 38/2; 97/2; 129/3; 154/1; 203/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;  
 C:Superfamily: collagen alpha 1(IV) chain



Query Watch	5.6%; Score 276.5; DB 2; Length 1870;
Best Local Similarity	22.7%; Presd. NO. 3.6e-06;
Matches 232; Conservative	75; Mismatches 358; Indels 359; Gaps 56;

  

QY	24	DQEVLGSH-SLQTPEEQGP-E-GVMG-----PWOWASCQPQGVGVQRSSRTC	71
DB	350	DEBGAAGHRDSASGEERPEADCKKGNPNSEPPTPKTAWAETSRP-----	397
QY	72	QLPTVOLHPSLLPRRPHREALLRGQGPRPOTSPETULPYRTOSRGGRGLRGPASH	131
DB	398	--PETEPGGPAKPPLPPPHRG---PAGNMGPGDYPD-----RGGP---PCKP	438
QY	132	LGREETEITRAARRRLRDPIKPMFGVGRVPALPLHNR-----RHRSPPRSLS	184
DB	439	PAPEDEDAWRQKKOSSSEI-----SLAVERARREREERMEOERRAACAA	486
QY	185	LTSRCGEAIPFTPR--AEFBSANGSPOTELPTTSLSVHTPSQAPHLSPATAQTEVAP	242
DB	487	EKLURDEKFADRLKAEEAAPASTPAPPFAVKELPAPPAPP--PASAPT---P	541
QY	243	RTRPAPLRHHRAQASGTE--PPSPTHSLGEG-----GFFRAS-PQRRRSSOGWASP	292
DB	542	EKEPEEAPAQAOSTTPGVAAAPTTLVSGGSTSTSSGSFEASPVFQLPKSKGPEPP	601
QY	293	OYAGREDDPPSVPR-----CRQOQGGP-WGTGGTHPCRPLEDPQHPGWLPLSNG	345
DB	602	EEV--PPPTTPVPVKEPKGDIGTTPQPSQGLGYKYOKSLPFPRQOOEQILLKQQQ	659
QY	346	PHASSLWSLF----APSSIFPRCSGEZQLRACSAQCPPEQPDPRALOCASFNSQEFMG	401
DB	660	QH-Q---WOHQQSAPPTVPSPQVPTLCGAVPAKAPB--PPPKAL-----YPG	705
QY	402	QLYWOPFTTEVOGSORCELNCRBRGF-F-YVRHTEKVQDGTLCPGADICVAGRCLSP	459
DB	706	ALGRPPMP-----PWNFDPRMMIPXY-----DRLUQ-----GR---P	738
QY	460	GCD-----GILGSG-----RRPDGCGVGCGDDSTCLRVSGLTDGGLGYOKILWI-	506
DB	739	PLDFYPPGVHPGLVPRSERSGLI--SSEPDRHAPAMLRERTGTPVDVKLAWGDVFT	796
QY	507	--PA-----GALRLQIALRPSNNYA-----LR	528
DB	797	ATPAEPRPLTSLRQAADDKMKRSETPPVPPPXYLASYPGFENCTGPPISRFPLE	856
QY	529	GPGRSIINGNAVDPGYSRAGTVFYRNRP--REEKGESLSAEGPTT-----	577
DB	857	EPGPREL--PW---PPGS---DEVAKIQTPPKPEKPEETAQTGTPEAGRKLPASRS	906
QY	578	-----QPVDVMTFOENPCVFQYVISSPPPILENPTPEPVP--	616
DB	907	GAGPPPPRESRTETWGPGRGSSRRGIPEPCA-----PPRAGTKKPPPTK	957
QY	617	--OLOPEILRV--EPLAPARPAR-TGTLQROVRIPQMAPPHPRTPLG--SPA-	666
DB	958	VEELPKPKLEQDETPKPKDPDKITKGL-----GGKETPPGNLSPAPLR	1007
QY	667	---YMKRVGHSCASCSCKGVYRWDFLCISR-----ESGELDER	703
DB	1008	RDYSERVQFTSC-----RGRGREYFARGRGFTYGGRGCAQANSAVTESFEMGV	1062
QY	704	SZAAGAR-----PPASBPCHGTTCPPPYWEAGEWTSRCSGPGTOHROLQCRBFGGG	758
DB	1063	EVGGQDTLLLEPALPARHCA-----RVQSMRKSPSGAG	1098
QY	759	SSVPPPERCGLFRP-NITOSQLRLCGHWGVGSPWSCSVRCRGORSR-----	808
DB	1099	SGAQKAA---RPMRVILWLTQLRLPH-----PRHSPPRSRSRTTSPILHRAPAF	1149
QY	809	RCVGNNGDEVSOECASG-----PPQ-----PPSREACDMGPCWTAFHSDWSKSYSP	856
DB	1150	TCPGVGESLPGEATSPPGRREAPPQVCWGSPSPAKSLAKPPPT-----GP	1197

Qy	857	EPPA 860
Db	1198	LPPS 1201
 RESULT 14		
S36152		
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu		
C:Species: Homo sapiens (man)		
C>Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000		
C:Accession: S36152		
R:Iris, F.U.M.; Bouquelaret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Ju		
Nature Genet. 3, 137-145, 1993		
A:Title: Dense AU clustering and a potential new member of the NKG2PB family with		
A:Reference number: S36152; MUID:93272029		
A:Accession: S36152		
A>Status: preliminary		
A:Molecule type: DNA		
A:Residues: 1-1872 <IRI>		
A:Cross-references: EMBL:Z15025		
A>Note: in the authors' translation residues 32-34 are shown after residue 4 and, con		
A:Note: the authors translated the codon AAT for residue 1000 as His		
C:Genetics:		
A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;		
C:Superfamily: collagen alpha 1(IV) chain		
 Query Match 5.6%; Score 276.5; DB 2; Length 1872;		
Best Local Similarity 22.7%; Pred No. 3 6e-06;		
Matches 232; Conservative 75; Mismatches 358; Indels 359; Gaps 56;		
Qy	24	DQVLSSH-SLOTTERGGQPE--GVWG-----PWQWASCSQPQCGVGVORRSVC 71
Db	351	DEGAEGHRDSQSASGERPEADGKKGNSPDPPTKTAWAETSRP----- 398
Qy	72	QLPTVQLHPSLPLPRPRRHPEALLPGQGPRPOTSPTLLPYRTQSRGGGLRGPAHS 131
Db	399	--PETEGPPAPKPPLPPPHRG---PAGNWGPDDYD-----RGGP--FCKP 439
Qy	132	LGREETQEIRAAERSRLRDPIKGMFGVGRVPFPALPLHRNR-----RUPSPPRSLS 184
Db	440	PAPDEEDAMQRKKSSSEI-----SLAVERRRREREERMQEERRAAC 487
Qy	185	LISRGEEAIPSPTR--AEFVSANGSPQTLPTELTVHTTSPQAEPSPETAQTEVAP 242
Db	488	EKLRLKEKFCAQPKRLKAEPAPAAPSTPAPPVAPVKELPAPPAPP--PASAPT--P 542
Qy	243	RTRAPLRHHPRQAACSTE--PPSPTHLEGG-----GFERASP--OPRRPSSQGWASP 292
Db	543	EKEEPEAQAPPAQSTTPPGVAAPTLVSGGSTTSISSGSFEASPVBPQLPSKEGEPPP 602
Qy	293	QVAGRDRPFPSVPR-----GRQQCGQP--WGTGTPHGRLPDFQHPCAMLPLLNSG 345
Db	603	EEV--PPPTTPPVKVBFKGDIGFTRPQPSQSLGYKYOKSLPPRFQROQQEQOLLKQQQ 660
Qy	346	PHASSLSLF----APSSPIRCSESDLRACSAQCPCPEQPDPRALQCAAFNSQEFMG 401
Db	661	QHQ---NQOHOOGSAPTPVPSPQPVTLCGVAPAKAPP--PPPKAL-----YPG 706
Qy	402	QLYWEPFTEVOGSRCELNCRPGFRF--YYVRHTEKVQDGTLCQPGADPICVAGRCISP 459
Db	707	ALGRPPMP-----PMNFDPRMMIPYV-----DPRLLQ-----GR--P 739
Qy	460	GCD----GILGSG-----RRPDCCGVCGDDSTCLVSGNLTDGCGPLYGOKILWI----- 506
Db	740	PLDYPPGVHFHSLVPRSDSRL--SSEPFDRAFAMLREGRTPVPDKLAWGVDFVT 797
Qy	507	--PA-----GALRLQIALRPPSNLYA-----LR 528
Db	798	ATPAEAPRLTSLRQAADDKDKMRSTPVPVPPPPPYLASYPGPENGNGTCGPPISRPLE 857
Qy	529	GPGRSIINGNWVDYPCSYRAGCTVFYRNRP--REEKGESLSAEGPT----- 577



Db 858 EPGRPPL---PW-----PPGS-----DEVAKIOTPPPKKEPKEETAQITGPEAGKILPASRS 907  
Qy 578 -----QFVDVIMFOENPGVEYQVVISPPPILENPTPEPVP-- 616  
Db 908 GAGPPPPRRSRTETRWGPRGSSRRGIPPEEFGA-----PPRRAGPIKKPPPTK 958  
Qy 617 --QLOPEILRV--EPPLAPAPRAR--TPGTLOQVRIPOMPAPPHPTPLG--SPAA--- 656  
Db 959 VEELPPPLBOGDETPKPPKPDPLKTKGKL-----GGPKETPPNGNLSAPRLR 1008  
Qy 667 ---YWKRVGHSASCSGKGVWRPFILCISR-----ESGELDER 703  
Db 1009 RDYSYRVGTSC-----RCGRGEYFARGRGFRGTGYGGRGCAQANSAVTESFEEMGV 1063  
Qy 704 SCAAGAR-----PPASPEPCHGTPCPYWEAGEWTSRSCPGTGHROLOCRQEFEGGG 758  
Db 1064 EVGOGDQTTLLPEALPARHGA-----RVQSMRKSPSGAG 1099  
Qy 759 SSVPERCGHLPAP-NITQSCQLRLCGHMEVGSQVRCRGORSR-----QV 808  
Db 1100 SGAQKQAA---RPMRVILWLTQLRLPH-----PREHSPRSRSTTSPTLHRAPARF 1150  
Qy 809 RCVGNNGDEVEQECASG-----PQO-----PPSREACDMGCPCTAWFSDMSKVSVP 856  
Db 1151 TCPGVGSSLPFGAISGPRRRREAPPQVCPGWSPPAKSLAPKKPT-----GP 1198  
Qy 857 EPPA 860  
Db 1199 LPPS 1202

RESULT 15  
Q0BE3  
BHLF1 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 23-Aug-1997  
C:Accession: A03742  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-43, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A:Reference number: A93065; MUID:85035713  
A:Accession: A03742  
A:Molecule type: DNA  
A:Residues: 1-660 <BAN>  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H  
Nature 310, 207-211, 1984  
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
A:Contents: annotation; protein coding region  
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52  
C:Superfamily: human herpesvirus 4 BHLF1 protein

Query Match 5.4%; Score 262; DB 1; Length 660;  
Best Local Similarity 22.4%; Pred. No. 6.9e-06;  
Matches 206; Conservative 51; Mismatches 293; Indels 368; Gaps 50;

Qy 35 TPTEGQGEVGMVQWVQWASCSQPCGVGVGVRSSRTCOLFTVQLHPSLPLPPRPPRHPEA 94  
Db 3 TPCQSARGP-----RTTLP-----HCPPLCLGAPDQQT 33  
Qy 95 LLPRGQGR-----POTSPTLPLYTOSRGSGGLRGFASHLGHRETEIRAARR 145  
Db 34 RLPPGWQRTAPTQVGLADASDELDQASGARGGGNRVG-----AGRGEGTTPAPSRQ 89  
Qy 146 SRLRDIKPGMFGYGVFPFALPHRRNRHPPRPSRSELSSIRGEEAIPSPTPRAEP-- 203  
Db 90 SRRTGPAEQADHAHNS-----PTGGCSDPORSPTRCAGYALGEGSAGLSGRGPRHPAF 144  
Qy 204 ---FSAN--GSPQT-----ELPTELSVHTPSQ--AEPLSPETAQTEVAP-----R 243  
Db 145 QVQWSARNPGCPRTWRRRSRGAQRGHPGAGQRPSTGGRPAAPGPGTTPAAGPGGGGA 204

Search completed: July 23, 2002, 21:06:30  
Job time: 2784 sec

Qy 244 TRPAPLRHHFRAQASGTEPPSPTHSLGEGGFFRASP---QPRRPS-----S 286  
Db 205 AVPSGATPHERGSGPADPFAA-----RLPPERQEPRLQDLAAQRCAPGPPPT 255  
Qy 287 QWASPOVAGRRDDPPPSVPRGRG-----QOGQGPWGTGCTPHGPRLEPDPQHP 335  
Db 256 RSGAAARTTHRRPPGCRSARNPGCPRTWRRRSRGAQRGHPGAGORPSGP--TGGRPAAP 314  
Qy 336 GAWLPLLSNGSPHASSLSLFASSPIP---RCSGESEQLKACSAQCPPEQDPFALQCA 392  
Db 315 GAPCTPAAPGPGGAA---VPSGATPHPERGSPADPFAA---ARLPPERQEPRLPQ-- 365  
Qy 393 AFNSQEFMGOLYQWEPTFVQSGORCELNCRPRGRFYVHTEKVKVQDGTLCQPGADICV 452  
Db 366 -----DLAAQRC-----PAGPPPT 381  
Qy 453 AGRCLS-----PCCDGLGRRRPGCGVCGGDDSTCLVSGNLDRGGPLGYQKILWI 506  
Db 382 SGAAARTTHRRPPGCP---HSARNP-GC-----PRTWRRSG--AQRGHP-- 420  
Qy 507 PAGALRLQIAQLRPPSSNYLALRGP--GGRSINGNWAVDPPGSYRAGGTVFRYNRPPREG 565  
Db 421 PPGAGQ-----RPS-----OPTGGR-----PAAPGAPGT-----PAAPG 449  
Qy 566 KGESLSA--EGPTTQPVVDVYMIFOENPGVYQYVISPPPILENPTPEPPVQLOPEILR 624  
Db 450 PGGGAAPVSGATPEP-----ERGS-----PADPFAAARLPPERQEPRL-PQDLA 493  
Qy 625 VEPPLAPAPPARTPGLQROVRIPOMPAPPHPTPLGSPAAYKXVGHASCSGKGV 584  
Db 494 AQRCAPGPPPTRSGAAARTTHRRP---PGCPR-----SARNPGCPR-T 533  
Qy 685 WRPIFLCISRESGEELDERSCAAGARP-----PASPEPCHGTGCPPYWEAGEWTSCSR 737  
Db 534 WR-----RRSGAQRGHPPPGAGQRPSTGGRPAAP--GAPCTPAAP----- 573  
Qy 738 SCGPG-----TQHRQLQCRQEFEGG-----CSSLVPERCGHLPRNITQSCQLRL 782  
Db 574 --GGGGAAPVPSGATPH-----PERGSGPADPFAAARLPPER---QEPRLPQ----- 615  
Qy 783 CGHWEVGSQVSWQSVRCGRGQRQRVRCVGNNGDEVSEQECASGPP-----Q 829  
Db 616 -----DLAAQRCAPGPPPTRSGAAARTTHRR 642  
Qy 830 PPS--REACDMGCPCTTAW 845  
Db 643 PPGCPRSRNPG-CPRTW 659

OM of: US-10-041-770-2 to: GenEmbl.\* out\_format : pfs  
Date: Jul 23, 2002 10:28 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 compugen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US10041770/runat\_22072002.153135.2679/app\_query.fasta\_1.947  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOPEXT=0.000 -CGAPOP=4.500  
-CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -START=1 -MATRIX=bicsum62 -TRANS=human40.cdi  
-DELETE=7.000 -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US10041770 -CGNL=1.5038  
-NCP=6 -ICP=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-10-041-770-2  
Query length: 877  
Database: GenEmbl.\*  
Database sequences: 1797656  
Database length: 187333701  
Search time (sec): 2782.860000

score\_list:

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
gb_pat:AX342636	+	4769.00	1381.91	1.4e-68	4230	AX342636 Sequence 33 from Pat
gb_pat:AL356356	+	3076.50	1139.91	3.9e-55	176550	AL356356 Human DNA sequence
gb_hg:AC092479	+	2870.50	769.90	1.6e-34	142023	AC092479 Mus musculus clone
gb_hg:AC093317	+	2851.50	762.87	4.1e-34	215810	AC093317 Mus musculus clone
gb_hg:AC053497	+	2070.50	597.80	6.2e-25	207815	AC053497 Homo sapiens chrom
gb_hg:AC053497	+	2014.50	581.90	4.8e-24	207815	AC053497 Homo sapiens chrom
gb_pat:AF217974	+	1231.00	379.63	8.8e-13	2070	AF217974 Homo sapiens clone PH
gb_pat:BC008840	+	935.50	294.52	4.8e-08	2703	BC008840 Homo sapiens, hypothe
gb_pat:ES5265	+	807.50	255.40	7.3e-06	5061	ES5265 Novel metalloprotease a
gb_pat:AX128433	+	806.50	254.87	7.8e-06	5353	AX128433 Sequence 2 from Paten
gb_pat:AX138851	+	797.50	254.43	8.3e-06	3312	AX138851 Sequence 15 from Pat
gb_pat:ES5273	+	796.50	254.14	8.6e-06	3312	ES5273 Novel metalloprotease d
gb_pat:AF163762	+	791.50	252.61	1.0e-05	3400	AF163762 Homo sapiens zinc met
gb_pat:AC023772	+	789.50	252.64	1.0e-05	2964	AC023772 Homo sapiens cDNA FL
gb_pat:AX327754	+	784.00	250.84	1.3e-05	3132	AX327754 Sequence 10 from Pat
gb_pat:AX327752	+	784.00	250.57	1.4e-05	3329	AX327752 Sequence 8 from Paten
gb_pat:AX319854	+	780.00	250.19	1.4e-05	2805	AX319854 Sequence 11 from Pat
gb_pat:AX327755	+	759.50	243.43	3.4e-05	3471	AX327755 Sequence 18 from Pat
gb_pat:AX111571	+	757.00	243.07	3.6e-05	3199	AX111571 Sequence 1 from Paten
gb_pat:AX327747	+	757.00	243.06	3.6e-05	3307	AX327747 Sequence 3 from Paten
gb_pat:AX338539	+	757.00	242.80	3.7e-05	3402	AX338539 Sequence 3 from Paten
gb_pat:AX327745	+	757.00	242.80	3.7e-05	3403	AX327745 Sequence 1 from Paten
gb_pat:AX338537	+	757.00	242.80	3.7e-05	3403	AX338537 Sequence 1 from Paten
gb_pat:AX319852	+	753.50	241.47	4.4e-05	3675	AX319852 Sequence 16 from Pat
gb_pat:AF314171	+	731.00	236.55	8.2e-05	4271	AF314171 Mus musculus papilin
gb_in:AF078161	+	731.00	230.25	0.0002	11008	AF078161 Manduca sexta lacuna
gb_in:AF205357	+	722.50	229.77	0.0002	7100	AF205357 Drosophila melanogast
gb_pat:AX111573	+	714.50	232.52	0.0001	2266	AX111573 Sequence 3 from Paten
gb_hg:AC014412	+	699.50	218.94	0.0019	86446	AC014412 Drosophila melanog
gb_in:AC009213	+	699.50	218.94	0.0019	182525	AC009213 Drosophila melanog
gb_in:AC009213	+	699.50	218.94	0.0019	182525	AC009213 Drosophila melanog
gb_in:AX319861	+	698.50	226.34	0.0003	3285	AX319861 Sequence 25 from Pat
gb_in:AF043121	+	695.50	223.47	0.0004	5199	AF043121 Haemochus contortus
gb_in:AF011177	+	666.00	219.01	0.0008	4067	AF011177 Homo sapiens mRNA for
gb_pat:HS250725	+	666.00	215.16	0.0013	5115	HS250725 Homo sapiens mRNA for
gb_pat:AX319860	+	636.50	209.08	0.0028	2937	AX319860 Sequence 32 from Pat
gb_pat:AX342635	+	626.50	206.39	0.0039	2830	AX342635 Sequence 24 from Pat
gb_pat:AX342636	+	601.00	200.01	0.0089	2410	AX342636 Sequence 35 from Pat
gb_pat:AX342638	+	598.50	196.87	0.0133	4185	AX342638 Homo sapiens mRNA for
gb_pat:HS305314	+	596.00	196.78	0.0135	3636	HS305314 Homo sapiens mRNA for
gb_pat:HS343125	+	596.00	196.78	0.0135	3636	HS343125 Homo sapiens mRNA for

gb\_pr:AY055376 + 593.50 195.04 0.0168 4597 1 AY055376 Homo sapiens von W  
gb\_pr:HS420810 + 590.50 194.50 0.0180 4278 1 AJ420810 Homo sapiens mRNA  
gb\_pr:AF414401 + 590.50 194.23 0.0187 4550 1 AF414401 Homo sapiens ADAMT  
gb\_pr:AB069698 + 590.50 193.86 0.0196 4950 1 AB069698 Homo sapiens VWF-C  
gb\_pr:AB037733 + 589.00 193.27 0.0211 5139 1 AB037733 Homo sapiens mRNA

seq\_name: gb\_pat:AX342636

seq\_documentation\_block:

LOCUS AX342636 4230 bp DNA linear PAT 12-JAN-2002  
DEFINITION Sequence 33 from Patent WO0198468.  
ACCESSION AX342636  
VERSION AX342636.1 GI:18152033  
KEYWORDS  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)  
Yue H., Elliott V.S., Gandhi A.R., Lal P., Au-Young J.,  
Tribouley C.M., Deleage A.M., Baughn M.R., Nguyen D.B., Lee E.A.,  
Hafalia A., Khan F.A., Wallia N.K., Yao M.G., Lu D.A., Patterson C.,  
Tang Y.T., Walsh R.T., Azimzal Y., Ramkumar J., Xu Y. and Reddy R.,  
Patent: WO 0198468-A 33 27-DEC-2001;  
Incyte Genomics, Inc. (US)

LOCATION/Qualifiers

1. 4230

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/note="Incyte ID No: 7604035CB1"

BASE COUNT 810 a 1411 c 1210 g 799 t

ORIGIN

alignment\_scores:

Quality: 4769.00 Length: 857  
Ratio: 5.584 Gaps: 0  
Percent Similarity: 99.650 Percent Identity: 99.650

alignment\_block:

US-10-041-770-2 x AX342636

Align seg 1/1 to: AX342636 from: 1 to: 4230

1 MetGluAnTrpThrGlyArgProTrpLeuTrpLeuLeuLeuSe 17

||||| 228 ATGGAGAACTGGACTGGCAGGCCCTGGCTGCTGCTGCTCTCTGTC 277

||||| 17 rLeuProGlnLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuG 34

||||| 278 CCTCCCTCAGCTCTGCTGGATCAGAGGTTGTTCTCCGACACTCTCTTC 327

||||| 34 InThrProThrGluGluGlnGlyProGluGlyValTrpGlyProTrp 50

||||| 328 AGACACCTACAGAGAGGCGGCCGAGGCGCCGAGGAGTGTCTGGGACCTTGG 377

||||| 51 ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgAr 67

||||| 378 GTCAGTGGGGCTTGTCTCCAGCCCTGGGGTGGGGTGGGAGCCAG 427

||||| 67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuProL 84

||||| 428 GAGCGGACATGTCAGCTCCCTACAGTGCAGTCCACCGAGTCTGCC 477

||||| 84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100

||||| 478 TCCCTCCCGCGCCCGCCAGACATCCAGAAAGCCCTCCCTCCCGGGCCAG 527

||||| 101 GlyProArgProGlnThrSerProGlnThrLeuProLeuTrpArgThrGl 117

||||| 528 GGTCCAGACGCCAGACTTCTCCAGAACCCCTCCCTGTACAGGACACA 577

||||| 117 nSerArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyA 134

1478 CTGTGAAGTAACTGCCGCCCGCTGGCTTCCGCTTCTATATCTCCGTCACA 1527  
 434 hzGluLysValGlnAspGlyThrLeuCysGlnProGlyAlaProAspIle 450  
 1528 CTGAAGAGTCCAGGATGGACCTGTGTGACCTGGAGCCCTGACATC 1577  
 451 CysValAlaGlyArgCysLeuSerProGlyCysAspGlyIleLeuGlyse 467  
 1578 TGTGTGGTGGACGCTGTGAGCCCGCTGTGATGGATCCTTGGCTC 1627  
 467 rGlyArgArgProAspGlyCysGlyValCysGlyValAspAspSerThrC 484  
 1628 TGGCAGGCTCTGTGATGGCTGTGAGTCTGTGGGGTGTGATCTTACCT 1677  
 484 yArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyr 500  
 1678 GTCCGCTTGTTCGGGGAACCTCACTGACCGAGGGGCCCTTGGGCTAT 1727  
 501 GlnLysIleLeuTyrIleProAlaGlyAlaLeuArgLeuGlnIleAlaG 517  
 1728 CAGAGATCTTGTGATTCACGCGGAGCCTTGGGCTCCAGATTGCCCA 1777  
 517 nLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgS 534  
 1778 GTCCGCGCTAGCTCACTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1827  
 534 erIleIleAsnGlyAsnTyrAlaValAspProGlySerTyrArgAla 550  
 1828 CCATCATCAATGGAACTGGGCTGTGGATCCCGCTGGGCTCTACAGGCC 1877  
 551 GlyGlyThrValPheArgTyrAsnArgProArgGlyGlyGlyGlyGly 567  
 1878 GGGGGGACGCTCTTCGATATTAACGCTCTCCAGGGAGGGGCAAGG 1927  
 567 yGluSerLeuSerAlaGluGlyProThrThrGlnProValAspValTyrM 584  
 1928 GGAGAGTCTGTGGCTGAAGGGCCCAACCCAGGCTGTGGATGCTATA 1977  
 584 etIlePheGlnGluAsnProGlyValPheTyrGlnTyrValIleSer 600  
 1978 TGATCTTCAGGAGGAAACCCAGGGGTTTTTATCAGTATGTCATCTCT 2027  
 601 SerProProIleLeuGlnAsnProThrProGluProProValProG 617  
 2028 TCACCTCTCCATCTTGAAGACCCCAACCCAGGCTGTGGGCTGTCCCA 2077  
 617 nLeuGlnProGluIleLeuArgValGluProProLeuAlaProAlaProA 634  
 2078 GCTTCAGCGGAGATTCAGGGTGGAGCCCACTTGTCCGGCACCCC 2127  
 634 rgProAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGln 650  
 2128 GCCCAGCGCGACCCAGGACCTCCAGGCTCAGGTGCGGATCCCCAG 2177  
 651 MetProAlaProProHisProArgThrProLeuGlySerProAlaAlaTy 667  
 2178 ATCCCGCGCGCCCATCCAGGACACCCCTGGGCTCTCCAGCTGGTA 2227  
 667 rTrpLysArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyV 684  
 2228 CTGAAACAGAGTGGACACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2277  
 684 alrTrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeu 700  
 2278 TCTGGCGCCCATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2327  
 701 AspGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluPr 717  
 2328 GATGAACGACGCTGTGCGGGGTGCGAGGCCCGCCAGCTCCCTGAAAC 2377  
 717 cCysHisGlyThrProCysProProTyrTrpGluAlaGlyGluTrpThrS 734  
 2378 CTGCCACGGCACCCCATGCCCCCATCTACTGGAGGCTGGCGAGTGGACAT 2427

578 GTCTCGGGGAAGGGTGGCCACTTCAGAGTCCCGCTTCCCACTAGGA 627  
 134 rgGluGluThrGlnGluIleArgAlaAlaArgArgSerArgLeuArgAsp 150  
 628 GAGAGGAGACCCAGGAGATTCGAGCGCCAGGAGGTCCCGGCTTCGAGAC 677  
 151 ProIleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuPr 167  
 678 CCCATCAGCCAGGAAATCTCGGTATGGGAGAGTGCCCTTTCATTTGCC 727  
 167 oLeuHisArgAsnArgArgHisProArgSerProProArgSerGluLeuS 184  
 728 ACTGCACCGGAACCGCAGCACCTTCGAGGCCACCCAGATCTGAGCTGT 777  
 184 erLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProArg 200  
 778 CCTGTATCTCTTAGAGGGGAGAGCCCTATTCGCTCCCTACTCCAGA 827  
 201 AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProTh 217  
 828 GCAGAGCATTTCTCGCAAGCGGAGCGCCCAAACTGAGCTCCCTCCAC 877  
 217 rGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG 234  
 878 AGAAGTGTCTGCTCCACACCCCTATCCCGCCAGGAGAACCTCTAAGCCCTG 927  
 234 luThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArg 250  
 928 AAACGTCTCAGACAGAGTGGCGCCCGCCAGAACAGGCTGCGCCCTACGG 977  
 251 HisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrHi 267  
 978 CATCACCCAGAGCCAGGCTCTGGCACAGAGCCCGCTCACCACAGCA 1027  
 267 sSerLeuGlyGluGlyGlyPheArgAlaSerProGlnProArgArgP 284  
 1028 CTCTTAGGAGAAGTGGCTCTTCCGTGATCCCTCAGCCACAGAGGC 1077  
 284 roSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArgArgProAsp 300  
 1078 CAAGTTCACAGGTTGGGCCAGTCCCGAGGTAGCAGGAGAGCGCCCTGAT 1127  
 301 ProPheProSerValProArgGlyGlyGlnGlnGlyGlnGlyProTr 317  
 1128 CCTTTTCTCGTCCCTCGGGGCGAGGCCAGCAGGCGCAAGGCGCTTG 1177  
 317 pGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspProGlnH 334  
 1178 GGGAAAGGGGGGAGTCTCTACGGGCGCGCTGGAGCTTACCTCAGC 1227  
 334 isProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSerSer 350  
 1228 ACCCGGCGCTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1277  
 351 LeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyGl 367  
 1278 CFTGTGAGGCTCTTGTCTCCAGTAGCCCTATTCAGAGTGTCTGGGA 1327  
 367 uSerGluGlnLeuArgAlaCysSerGlnAlaProCysProProGluGlnP 384  
 1328 GAGTGAACAGCTAAGAGCCTGACGAGCCGCGCCCTGCCCCCTGAGCAGC 1377  
 384 roAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMet 400  
 1378 CAGACCCCGGGCCCTGCGAGTGGCGAGCCCTTAACCTCCAGGAATTCATG 1427  
 401 GlyGlnLeuTyrGlnTrpGluProPheThrGluValGlnGlySerGlnAr 417  
 1428 GGGCAGCTGTATCAGTGGGAGCCCTTCACTGAAGTCCAGGGCTCCAGCG 1477  
 417 gCysGluLeuAsnCysArgProArgGlyPheArgPheTyrValArgHisT 434

750 SerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCys 750  
751 ArgGlnGluPheGlyGlyGlySerSerValProGluArgCysGly 767  
767 yHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysG 784  
784 lYHisTrpGluValGlySerProTrpSerGlnCysSerValArgCysGly 800  
801 ArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGly 817  
817 uValSerGluGlnGluCysAlaSerGlyProGlnProProSerArgG 834  
834 luAlaCysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTrp 850  
851 SerSerLysValSerProGlu 857  
2428 CCTGCAGCCGCTCTGTGGCCCGCCGACCCAGCAGCCAGCTGCAGTGC 2477  
2478 CGGCAGGAATTTGGGGGGGTGCTCTCGGTGCCCGCCGAGCGCTGTGG 2527  
2528 ACATCTCCCCCGCCCAACATCACCCAGCTTGCCAGCTGCGCCTCTGTG 2577  
2578 GCCATTGGGAAGTTGCTCTCCTTGGAGCCAGTGCTCCGTCGGTGCAGC 2627  
2628 CGGGGCCAGAGAGCCGGCAGGTTCGCTGTGTTGGGACACACGGTGATGA 2677  
2678 AGTGACGAGCAGGAGCTGTGCTCAGGCCCCCCACACAGCCCCCAGCAGAG 2727  
2728 AGGCCTGTGACATGGGGCCCTGTACTACTGCTGTGTTCCACAGCGACTGG 2777  
2778 AGCTCCCAAGTGTCTCAGCCGAG 2798